# ABSTRACT

Title of Dissertation: **THE USE OF THE DOMESTIC DOG (CANIS FAMILIARIS) AS A COMPARATIVE MODEL FOR SPEECH PERCEPTION**

*Amritha Mallikarjun, Doctor of Philosophy, 2020*

Dissertation directed by: Professor Rochelle S. Newman
Department of Hearing and Speech Sciences

Animals have long been used as comparative models for adult human speech perception. However, few animal models have been used to explore developmental speech perception questions. This dissertation encourages the use of domestic dogs as a behavioral model for speech perception processes. Specifically, dog models are suggested for questions about 1) the role and function of underlying processes responsible for different aspects of speech perception, and 2) the effect of language experience on speech perception processes.

Chapters 2, 3, and 4 examined the contributions of auditory, attention, and linguistic processing skills to infants' difficulties understanding speech in noise. It is not known why infants have more difficulties perceiving speech in noise, especially single-talker noise, than adults. Understanding speech in noise relies on infants'

auditory, attention, and linguistic processes. It is methodologically difficult to isolate these systems' contributions when testing infants. To tease apart these systems, I compared dogs' name recognition in nine- and single-talker background noise to that of infants. These studies suggest that attentional processes play a large role in infants' difficulties in understanding speech in noise.

Chapter 5 explored the reasons behind infants' shift from a preference for vowel information (*vowel bias*) to consonant information (*consonant bias*) in word identification. This shift may occur due to language exposure, or possessing a particular lexicon size and structure. To better understand the linguistic exposure necessary for consonant bias development, I tested dogs, who have long-term linguistic exposure and a minimal vocabulary. Dogs demonstrated a vowel bias rather than a consonant bias; this suggests that a small lexicon and regular linguistic exposure, plus mature auditory processing, do not lead to consonant bias emergence.

Overall, these chapters suggest that dog models can be useful for broad questions about systems underlying speech perception and about the role of language exposure in the development of certain speech perception processes. However, the studies faced limitations due to a lack of knowledge about dogs' underlying cognitive systems and linguistic exposure. More fundamental research is necessary to characterize dogs' linguistic exposure and to understand their auditory, attentional, and linguistic processes to ask more specific comparative research questions.

# THE USE OF THE DOMESTIC DOG (CANIS FAMILIARIS) AS A COMPARATIVE MODEL FOR SPEECH PERCEPTION

by

Amritha Mallikarjun

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Rochelle S. Newman, Chair
Professor Jared M. Novick
Professor Jan Edwards
Professor Nan Bernstein Ratner
Professor Robert J. Dooling, Graduate Dean's Representative

# Acknowledgements

It takes a whole village to raise a dissertation – I'm pretty sure that's how the saying goes. Without all of the support from my advisor, postdocs, Ph.D. students, undergraduate research assistants, wonderful people in the community who have supported our lab, my friends in and out of academia, and my family, this would have been impossible.

First, a big thank you to my advisor Rochelle Newman. Rochelle's dedication to helping me become the best scientist I can be has been unwavering. Her support and belief in me, and all of her students, is made abundantly clear in the fact that Rochelle was willing to take a chance on starting a completely new research lab with a completely new species (that had never been tested before at Maryland, so without even a university precedent to rely on!). Whenever I encountered research speedbumps, I could count on Rochelle to help me through it with her signature straightforward assessments of the problem and a definitive plan to move forward.

On a different, non-academic note, Rochelle has introduced me to the wonderful world of sighthounds through her wonderful and hilarious dogs, Aela and Beren. At some point in my life when I own 7 whippets, you can thank Rochelle.

In sum, thank you for everything, Rochelle. I am immeasurably grateful for everything I have learned from you.

Thank you to all of the professors who have helped me throughout my time at the University of Maryland. Thank you to my committee members, Jared Novick, Jan Edwards, Bob Dooling, and Nan Ratner, for your invaluable advice and support.

Thank you to Yi Ting Huang, who helped me see that my wide variety of research interests could be an asset.

Thank you to Pam Komarek, an absolute pillar of support and an organizational wizard. You are the heart of the NACS program.

The next incredibly important thank you goes to my research partner and double-lab manager of both LDEV and the dog lab, Emily Shroads. From the inception of the Canine Language Perception Lab, Emily was the best partner I could ask for. We worked through all of the dog lab's early difficulties and growing pains together to make it into the productive research program it is today – and Emily did it all while simultaneously keeping the infant lab running smoothly. Emily provided incredibly valuable input on all the studies in this dissertation – without her, this dissertation would have been impossible. Thank you so much for your ideas, your support, and your unflinching belief in the potential of the dog studies.

Thank you so much to all of the Ph.D. students in the Language Development Lab over the years. We have laughed together, cried together, played Codenames together, and wore dinosaur hats together. An enormous thank you to Melissa Stockbridge for her friendship and support throughout grad school. You are an incredible person and I am so glad to know you. Thank you to Chris Heffner, my big brother, for all of the laughs and friendship. I know you will be an awesome professor.  A special thank you to Brittany Jaekel for sending me your dissertation proposal and giving me advice – it was invaluable. You are an amazing writer. Thank you to Erika Exton and Maddy Buntrock for inspiring me with your creative research ideas – I can't wait to see what you guys do next.

A special thank you to research assistant and postbac Rebecca Wolf, who provided tremendous emotional support and helped me realize that doing these dog studies was worth it. Many times I popped down to your apartment upset about my research or my thesis and we would watch some ridiculous reality TV, pet Sherman and the cats, and forgot about it all for a while. I hope that you will one day get a chance to own a border collie, and that you will teach him to do all sorts of amazing tricks.

Thank you to all the research assistants who have helped puppysit, schedule, soothe worried pups, and take cute photos for the Instagram: Chris Taho, Michaela Elm, Emma Peterson, Jillian Weinman, Christina Tyner, Jennifer Lee, Eva Johnson, Nicolette Contella, Matthew Tiberino, Ferddy Gedeon, Tilly Cornblatt, and Leila Hagopian. A special thank you to the very first research assistant for dog lab, Laura Grant – it has been an absolute pleasure to work with you and I know that you will be an incredible veterinarian.

Thank you to all the graduate assistants and postbacs who have helped with these projects. Early on, Lauren Steedman and Kelly Cavanaugh went through all our trial and errors with early studies to help the lab understand how to code dogs, and went on to train other research assistants on best coding practices. Kelly Puyear, Tiara Booth, Nada Babaa, Daniella Otarola, Mandy Giordano, and Veeda Baradar have all scheduled and coded countless dogs and helped me discover new research questions, new methodological improvements, and have graciously helped clean up various dog-related messes in the testing room, even though that is very outside your job

description. You are absolute stars and I know you will be extremely successful in whatever you decide to pursue.

Thank you to the LDEV postdocs Dr. Lucy Erickson, Dr. Katie Van Holzen, and Dr. Karen Mulak for all your valuable advice during this process. I'm so glad I got to learn from all of you.

Thank you to all of my friends that have listened to me go on long, excited rants about my dog research, and have supported me when I felt worried or stuck. An especially big thank you to my friend of over ten years, Jimmy Bartholomew, for listening to long rants about my chapters and helping me talk through my logic. Also, thank you to Jimmy for editing this acknowledgements page to make sure I didn't leave anyone out. My regular phone calls with you genuinely kept me sane during the breakneck cycles of writing and editing.

Thank you to my great friend and fellow graduate student Hannalore Gerling-Dunsmore for all of your support and for our great conversations on topics far and wide – it's so wonderful to have a friend who fundamentally understands how my brain works.

Thank you to Kelly Puyear for graciously offering up her Marie-Kondo-ing skills to not only the lab but also my closet. Thanks to you, I saved tons of time each morning when I got ready because I could actually find the things I was looking for.

Thank you to Elaine and AJ for your guidance and support throughout my time as a graduate student.

A brief thank you to my cats, Freyja and Kyrie, for their emotional support. They can't read, but I'm sure they will appreciate this acknowledgement. Well, Kyrie may have actually wanted co-authorship of this dissertation, but I didn't think her contributions warranted that.

An important thank you to my mother, Kamakshi Mallikarjun, and my father, Ramesh Mallikarjun. Mom - Your endless support and encouragement has been incredibly important to me. Thank you for being as excited about my dog research as I am – I'm so happy that you now love dogs and animals like I do. Dad – Thank you for always believing in me. I'm so thankful that you shared your own experience getting a Ph.D. with me when I got frustrated and worried, so I knew I wasn't alone. Lastly, thank you both for entertaining Freyja and Kyrie when they were making it difficult to concentrate on my "dessert station"– we all know playing with them is a full time job. Without both of you, this dissertation would not have been possible.

Thank you to my late grandfather, Professor R. Mallikarjunan, who I call Malli Thatha. You have been a champion of my academic pursuits since I was old enough to walk. I always wanted you to see me get a Ph.D. – I hope you would be proud.

Thank you to Hero Dogs Inc., an amazing local organization that raises and places service dogs and other skilled dogs. Thank you for bringing your dogs to the lab and for your enthusiasm for our work. I'm so grateful for all that I have learned while volunteering for Hero Dogs, and all the great advice I have gotten from the wonderful staff, trainers, and puppy raisers.

Lastly, thank you to all the dog owners that brought your dogs in to the lab – especially those of you that came in in the very beginning and helped us pilot test our studies (shout out to my friend Natalie and her dog Theo, the very first dog that every participated!). And lastly, a big thank you to all the dogs – you guys really made the magic happen.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Domestic dogs as a comparative behavioral model for speech perception in developmental populations

## *Overview*

The use of the domestic dog as a comparative behavioral model in psychology and neuroscience has dramatically increased in recent years. Dogs' domestication and evolution alongside that of humans led to a communicative, responsive animal that serves as an excellent model for comparative cognition studies (Andics & Miklósi, 2018). Dogs have primarily been used in comparison with adults, and most often in social cognition studies (e.g., Albuquerque et al., 2016; Cuaya, Hernandez-Perez, & Concha, 2016; Gácsi, Miklód, Varga, Topál, & Csányi, 2004; Huber, Racca, Scaf, Virányi, & Range, 2013; McKinley & Sambrook, 2000; Pilley & Reid, 2011; Racca et al., 2010; Range & Virányi, 2013; Wobber, Hare, Koler-Matznick, Wrangham, & Tomasello, 2009). This dissertation discusses how dogs could be useful comparative models specifically for developmental speech perception. Comparisons between humans and dogs are especially worthwhile for questions surrounding 1) how experience with and exposure to language affects speech perception, and 2) the identification of underlying mechanisms primarily responsible for different aspects of speech perception.

I will first discuss speech perception in adults and infants, and identify some challenges of studying speech perception with humans alone. I will then discuss several other animal models that have been previously used to study questions in speech perception, and how the dog model provides a useful non-invasive way to study some of these questions. I will then evaluate the potential advantages and disadvantages of a dog model for questions in developmental speech perception using the behavioral model evaluation framework from Holmes and Austad (1995).

### ***What is speech perception?***

Speech perception is characterized by the ability to perceive auditory input as "a sequence of meaningful linguistic representations", as stated by Gervain and Werker (2008, p. 1149). While adult humans can generally perceive speech without difficulty, speech perception is a complex task. Listeners must be flexible enough to recognize that a change in voice, intonation, or accent does not change a specific word's reference; however, listeners' representations must also be fine-grained enough to determine if a novel word is uttered. This is a distinctly difficult challenge given the incredible variation that can occur in the pronunciation of a single word. For example, the acoustic properties of a familiar word can change depending on characteristics of the speaker (i.e., speaker's sex, dialect, and emotional valence) and the context in which the word is heard (the other sounds surrounding the word). This problem is known as *lack of invariance,* and it is a classic issue in the speech perception literature (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). To contend with this variation, adults rely on their mature auditory system as well as their linguistic knowledge to recognize words across a variety of scenarios.

Variation in the speech signal itself is not the only complicating factor in word recognition. Listeners must also contend with the presence of distractors in the background (i.e., visual distractions like images on a television, or auditory distractions like other conversations or environmental noise). Comprehending speech of interest or relevance while ignoring background distractions requires mature attentional processes in addition to the auditory and linguistic processes necessary to contend with the acoustic signal.

Despite the complexities of speech perception, infants can accurately process speech, albeit with more difficulty than adults. Infants can do this even though they have limited experience with human speech, and their systems underlying speech perception (attention, audition, and linguistic processing) are not fully mature (Eimas, 1996, for lingustic system; Gomes, 2000, for attention system; Werner, 2007, for auditory system). It is difficult to address questions about linguistic experience and the underlying contribution of the auditory, attention, and linguistic systems to different speech perception tasks using infant subjects alone because 1) these systems are all in the process of developing (which makes it difficult to determine the underlying contributions via comparison of infants at different ages) and 2) there is little variation in the functioning of these systems in typically-developing infants at a specific age. Infant performance can be compared to adult performance; however, infants are in the process of developing all the systems involved in speech perception, and adults have fully mature systems and much more experience with language. These differences make it difficult to determine which systems and what type of experience are primarily responsible for infants' difficulties in certain speech perception tasks.

## *Prior models of speech perception*

Studying perception of speech in non-human animals can shed light on questions that would otherwise be impossible to test in humans alone. Prior animal models of speech perception have primarily been used as comparisons with adult humans. These models are selected either for analogies to human systems (e.g., selecting a zebra finch model for artificial grammar learning due to their regularized song system that shares similarities with the way human language is organized) or potential for homology (e.g., studying shared underlying processes, like the mammalian auditory system, to identify evolutionary precursors for aspects of language processing) (Kluender, Lotto, & Holt, 2006). Animal models are often used to examine whether certain speech perception processes are human-specific, or whether these processes are a result of more general cognitive processes. For example, chinchillas were used to assess whether animals could form categories for speech sounds. Chinchillas were able to learn categories for /d/ and /t/, and along a voice-onset-time continuum from /d/ to t/, they categorized the intermediary sounds similarly to adult humans (see Figure 1) (Kuhl & Miller, 1978) . This study provided evidence for the idea that phonetic categorization is not a human-specific ability, as the chinchillas must be utilizing general auditory and cognitive processes to learn these categories.

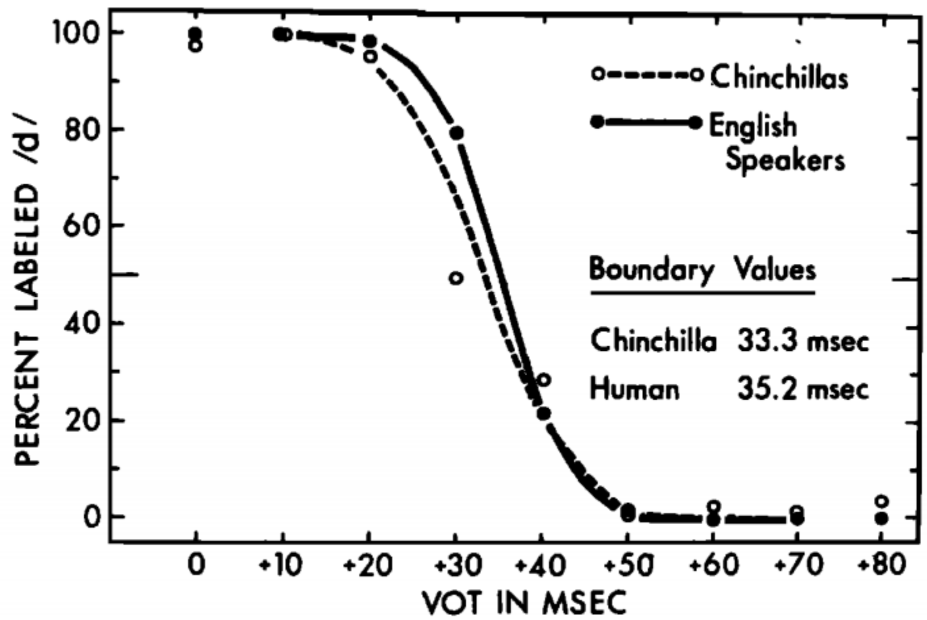


Figure 1: A graph from Kuhl & Miller, 1978, showing the mean percentage of /d/ responses by humans and chinchillas for intermediary sounds along a VOT continuum. Chinchillas and humans categorize these sounds similarly.

There are few animal models used specifically to examine typical developmental speech perception, and these models generally do not move beyond the question of whether a particular developmental speech perception ability is human-specific or not. For example, it is known that infants can distinguish early on between their mother's voice and other voices (Mehler, Bertoncini, Barriere, & Jassik Gerschenfeld, 1978) and that, as they gain more linguistic experience, they can distinguish between more voices (Johnson, Westrek, Nazzi, & Cutler, 2011; Vouloumanos, Hauser, Werker, & Martin, 2010). In a face-voice matching paradigm, several animals have been shown to distinguish between familiar and unfamiliar human voices (McComb, Shannon, Sayialel, & Moss, 2014; Saito & Shinozuka, 2013; Wascher, Szipl, Boeckle, & Wilkinson, 2012) and horses and Rhesus macaques can even distinguish between two familiar voices (Proops & Mccomb, 2012, for horses; Sliwa, Duhamel, Pascalis, & Wirth, 2011, for Rhesus macaques). While these results suggest that the ability to distinguish between human voices does not rely on human-specific processes, they cannot address the role of linguistic experience in infants' ability to identify voices.

The few studies that have moved beyond the question of human specificity often use models that are generally expensive and difficult to maintain (e.g., great apes, Lyn & Savage-Rumbaugh, 2000). Several studies have compared the ability of great apes to learn word forms and concepts to that of human infants, which has led to interesting findings about the aspects of speech perception that may not stem from the shared underlying processing of great apes and humans, and how much language can be learned through experience alone (see Ristau & Robbins, 1982, for an extensive review of great ape language-learning studies). However, there are many practical difficulties with studying great apes, ranging from financial limitations to ethical considerations in housing, enrichment, and experimentation (see Clark, 2011). An ideal developmental model of speech perception would allow researchers to study speech perception questions beyond that of human specificity, like the great apes, but the model would be more practically accessible for scientists to study.

### *Prior use of dogs as a comparative speech perception model*

Dogs have been used for more than a decade in language research to test questions related to human-specific language skills (Andics et al., 2016; Miklösi, Polgárdi, Topál, & Csányi, 1998; Ratcliffe & Reby, 2014; van der Zee, Zulch, & Mills, 2012). Many of these studies have explored the question of what aspects of speech perception are human-specific, and what aspects are derived from more general cognitive processes. For example, Rico, a German border collie, was shown to know labels for over 200 objects and could use *mutual exclusivity*, the idea that objects generally have only one name, to learn names for new objects (Kaminski, Call, & Fischer, 2004). When Rico was asked "*Wo ist der* (where is the) *blicket*", and the name *blicket* was an unfamiliar one, he would select a novel object over a known, previously named object. Young children also use mutual exclusivity to learn the names of new objects (Markman & Wachtel, 1988; Mather & Plunkett, 2011; Merriman, Bowman, & MacWhinney, 1989). The fact that dogs can show this same behavioral pattern suggests that the underpinnings of this skill may not be specific to language, nor dependent on cognitive skills unique to humans. More generally, understanding whether an ability is common to both canines and humans, or unique to one, can tell us a great deal about the underlying cognitive skills that allow for that behavior.

### *Assessing dogs as a model*

To address why testing dogs in addition to human subjects adds value to speech perception research, it is important to discuss the reasoning behind the selection of particular animal models to address specific research questions. In choosing a model animal for behavioral research, it is necessary to select an animal 1) that displays a particular behavior or trait of interest, or *specificity*; 2) where it is logistically possible and/or cost-effective to test this animal, or *feasibility;* and 3) where insights about the animal's behavior can lead to greater understanding about that behavior in humans, or *generality* (Holmes & Austad, 1995). Below, I use these model selection criteria to discuss the benefits and drawbacks of dogs as a model organism for speech perception research.

### *Specificity: Dogs attend to speech and learn word forms*

To use dogs as a model for developmental speech perception, they must demonstrate that they can perceive speech. Dogs' auditory systems can detect frequencies in the human speech range (Strain, 2011), so it is clear that they can hear human speech. In addition to their auditory speech perception capabilities, dogs also attend to human speech. Dogs also have evolved alongside humans to be particularly attentive to human behaviors and speech without any prior training (Hare, Brown, Williamson, & Tomasello, 2002). Both adult dogs (Horowitz & Bekoff, 2007) and young puppies (Ben-Aderet, Gallego-Abenza, Reby, & Mathevon, 2017) will treat vocalizations from humans as attention getters, and will orient their gaze towards the source of the vocalization. Dogs have also been shown to maintain their gaze to a source of audio if the audio is particularly interesting; for example, dogs will look longer when loudspeakers play their own name as opposed to another dog's name (Mallikarjun, Shroads, & Newman, 2019).

In addition to paying attention to human speech, dogs also have the ability to learn word forms and later recognize them. Work with several different individual dogs has suggested that some dogs may acquire vocabularies that are similar in size to those of young children (Griebel & Oller, 2012; Kaminski et al., 2004; Pilley & Reid, 2011); however, even dogs without special linguistic training have been shown to learn several different words. Pet dogs can recognize several commands, even at a young age (Kutsumi, Nagasawa, Ohta, & Ohtani, 2012). Additionally, pet dogs can distinguish between previously learned words, like their name, and unfamiliar words, even if they are said in the same intonation pattern (Mallikarjun et al., 2019). Together, these studies suggest that dogs demonstrate the ability to perceive speech.

### *Feasibility: How practical is it to test dogs?*

The feasibility requirement of an animal model states that a model should be accessible and cost-effective to use (Holmes & Austad, 1995). The domestic dog is an incredibly accessible and inexpensive behavioral model due to the fact that researchers can test local pet dogs brought in by their owners. Dogs do not need to be purchased, housed, or cared for on site, which greatly lowers cost and labor. Dogs are an extremely common household pet; in the United States, there are 89.7 million pet dogs (Springer, 2018). Additionally, people in the community are willing to bring their pet dogs in to participate in studies. Despite the fact that both the state of Maryland and the District of Columbia were amongst the states with the fewest pet-owning households (San Filippo, 2018), our lab has tested 723 dogs over three years. Together, these statistics suggest that there would be no shortage of dog participants for studies.

In terms of equipment, the necessary behavioral testing setup for comparative speech perception research would be similar to that of an infant lab. If an infant lab was to start comparative dog research, they likely would not have to purchase any additional testing equipment, and the costs would be limited to dog toys, treats, and cleaning supplies to ensure that dogs cannot smell prior participants in the testing space and that people who have dog allergies can safely use the space.

Dogs can often be used as models of speech perception with minimal to no prior training by experimenters due to natural variation in dogs' exposure to language and the number of words dogs recognize. Most model species would require an artificial training period in which they are exposed to particular stimuli, and then are tested on those stimuli at a later point in time (Bouchon & Toro, 2019; Toro & Trobalón, 2005). Dogs that live with humans are naturally exposed to words, such as their own name and several commands, from an early point in life. They hear these words in many different situations and from different people. This means that dogs would not necessarily require a training period during a study to familiarize them with a target word (Mallikarjun et al., 2019).

### *Generality: What can we learn about infant behavior from dog behavior?*

Dog models can provide a high level of generality such that they are informative about infant behavior in two different ways. First, dogs can be tested in paradigms that are very similar to infant paradigms, which allows for more direct comparisons of dog and infant experimental results. Second, when using a dog model for a speech perception study, the similarities and differences between the infant and dog systems underlying that speech perception task should be well understood; this allows the dogs' performance in that speech perception task to be used to better explain infant speech perception. I will first discuss paradigms, particularly the Headturn Preference Procedure, that can be used with minimal modifications for dogs and infants. Then, I will discuss the similarities and differences in the dog and infant linguistic, attention, and auditory systems in detail.

**Behavioral infant and dog testing paradigms.**

Several infant methods can be used to test dogs with very few changes; this allows for direct comparisons between infant and dog results in speech perception tasks (preferential-looking paradigm, Albuquerque et al., 2016; expectancy violation paradigm, Kundey, de Los Reyes, Taglang, Baruch, & German, 2010; the Headturn Preference Procedure, Mallikarjun, Shroads, & Newman, 2019). Unlike many other model animals, dogs are cooperative with humans and generally trainable, which eliminates the need for experimental manipulations involving food or liquid restriction. Dogs can be motivated by both social interactions as well as food (without the need for deprivation) and often their interest in humans and human vocal productions removes the need for food motivation. Here, I will discuss the commonly used developmental paradigm known as the Headturn Preference Procedure, which has been used in speech perception studies with both infants and dogs.

The Headturn Preference Procedure (HPP) is an infant experimental paradigm that measures infants' attention (as indicated by looking time) to auditory stimuli (Kemler Nelson et al., 1995). By comparing listening times to different types of auditory stimuli, it is possible to answer questions about infants' understanding about properties of their language. For example, 9-month-old infants listen longer to lists of words containing high-probability phonetic patterns in English rather than lists of words with low-probability phonetic patterns; this demonstrates that by 9 months of age, infants are sensitive to the frequencies of native language sound patterns (Jusczyk, Luce, & Charles-Luce, 1994). Moreover, infants typically find auditory stimuli more interesting when they match infants' expectations and prior experiences, as long as the items have not become so frequent as to be boring. By measuring infant preferences, we also gain an understanding of what the infant has already learned to expect about their language. For example, infants have been shown to listen longer to speech in their own language rather than a foreign language (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993) and to familiar words rather than unfamiliar words (Hallé & Boysson-Bardies, 1994).

HPP has also been proven useful in studying dogs' listening patterns; several studies in our lab have utilized HPP to show that dogs listen longer to their name over other dogs' names. Dogs can be tested in nearly the same way as infants, using the same apparatus (see Figure 2), the same stimuli, and almost exactly the same procedure. These similarities in testing allow us to directly compare results from dogs to results from infants. When researchers take advantage of dogs' gaze-following capabilities and their high level of interest in humans, many paradigms originally designed for infants can be utilized to test dogs in speech perception experiments.

Figure 2: On the left is a still image of an infant participating in a study using Headturn Preference Procedure, from Mind in the Making: Experiments in Children's Learning (Galinsky, 2011). On the right is an image of a dog participating in a study using Headturn Preference Procedure.

### *Comparing systems underlying speech perception.*

When using a nonhuman comparative model with the goal of shedding light on human behavior, like speech perception, the extent to which the model shares similar functional behaviors and similar underlying systems with humans determines the extent to which conclusions can be drawn about human behavior from the model's behavior. As such, to make claims about the mechanisms underlying infant speech perception from the results of dog studies, it is necessary to compare the underlying systems responsible for speech perception in both dogs and infants. In the sections below, I compare infant and dog linguistic, attentional, and auditory processing. Knowledge of the relationship between the infant and dog systems allows researchers to ask more precise questions of the model with more generalizability from dog to infant.

#### **Linguistic system.**

The *linguistic system* refers to the sets of underlying functions that govern language development and language processing. In this section, I will discuss and compare the development of linguistic abilities in infants and dogs. Typical infant linguistic development is rapid and generally predictable over time. Infants will alert to sounds shortly after birth (Morrongiello & Clifton, 1984), can recognize their own name by 4.5 months (Mandel, Jusczyk, & Pisoni, 1995), begin babbling by about 6 months (Oller, 2000), and produce their first words around 10-12 months (Boysson-Bardies & Vihman, 1991). From 18-22 months, infants' vocabulary grows exponentially (McMurray, 2007), and they begin to produce two-word utterances and follow two-word commands (Coplan, 1993).

While animals possess some of the linguistic capabilities of humans, they fail to demonstrate the vast communicative abilities of human language. Great apes (Patterson & Cohn, 1990; Savage-Rumbaugh, 1986), dolphins (e.g. Herman, 1986), and dogs (e.g. Pilley, 2013; Pilley & Reid, 2011) have all previously been taught to understand sets of words or symbols that represent different objects and actions, but they fail to use these words in the generative manner that humans do.

Dogs in general do not possess many of the characteristics of language in their own communication (Cohen & Fox, 1976). While domestic dogs have greater variability in their communicative vocalizations than their ancestors, wolves (Feddersen-Petersen, 2000), dogs still display a limited range of communicative vocalizations. Dogs use these vocalizations to communicate with other dogs (e.g., communication of desire to play, Faragó, Pongrácz, Range, Virányi, & Miklósi, 2010; agonistic growling to prevent other dogs from stealing food, Faragó, Townsend, & Range, 2014); however, there is no evidence of domestic dogs generating novel sounds for communication with either conspecifics or humans. As such, while dogs have the ability to communicate vocally, they are presumed not to have a human-like linguistic system. Their ability to learn certain aspects of human language must rely on skills other than a linguistic system, like dogs' auditory system, memory, and attention. By comparing dog and infant performance in linguistic tasks, it is possible to assess whether the underlying systems responsible for these tasks are human-specific, or whether they rely on more general mechanisms that mammals possess and can utilize for linguistic-like functions.

This section will discuss the similarities and differences in infant and dog language exposure, sound organization, or *phonology*, as well as their word-form learning and word-object mapping abilities. While the study of linguistics also includes aspects such as syntax and production, a domestic dog model is not well-suited for these types of questions, as they have not shown any significant capacity for syntactic processing and do not have a wide range of vocal productions; as such, these aspects will not be discussed in this paper.

***Language exposure.***

Infants' early language exposure has a large effect on their eventual language skills (Hart & Risley, 1995). The quantity and quality of language that infants hear has been shown to affect their eventual vocabulary development (Newman, Rowe, & Bernstein Ratner, 2016; Rowe, 2008; Weisleder & Fernald, 2013), as well as many other linguistic and cognitive skills (e.g., grammatical complexity, Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002; reading skills, Rodriguez & Tamis-Lemonda, 2011). While there is a great deal of research on infants' language exposure, not much is known about dogs' linguistic exposure or how it compares to that of infants. This impairs researchers' ability to use a dog model to study the effect of language exposure on the emergence of specific speech perception abilities. Further research is necessary to examine the quality and quantity of speech directed to dogs and how it is similar to and different from speech directed to infants. This could be done using a recording device in the home, similar to the manner in which this data was acquired for infants (Gilkerson & Richards, 2009). Better understanding of the speech directed to dogs would allow for more precise comparative speech perception studies examining the effect of language exposure.

***Phonological development.***

Phonology is the aspect of linguistics that deals with the organization and structure of sounds in spoken languages. Infants in approximately the first six to seven months of their life are very adept at discriminating between speech sounds. They not only successfully discriminate between sounds present in their own language, but also can discriminate between non-native sounds (Aslin, Pisoni, Hennessy, & Perey, 1981; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984). Through exposure to their native language, infants slowly begin to tune specifically to the sounds of their native language. After 12 months, infants stop distinguishing between sounds that are not contrastive in their native language (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984); this means infants actually discriminate between *fewer* contrasts than they did previously. When infants focus on their native phonology, they begin to build phonetic representations, which are abstract linguistic representations of sound categories.

There are few studies that have explored dogs' perception of human speech at the phonological level. One early study demonstrated that young dogs who have not had much language exposure show categorical labeling of consonants, or the labeling of distinct consonant categories along a continuum (Adams, Molfese, & Betz, 1987). This is in line with evidence showing many other vertebrate animals, including chinchillas (Kuhl & Miller, 1978), guinea pigs (Sharma & Dorman, 1999), rats (Reed, Howell, Sackin, Pizzimenti, & Rosen, 2003), and rhesus monkeys (Morse, Molfese, Laughlin, Linnville, & Wetzel, 1987), naturally differentiate between human speech sound categories without prior training. What is unknown is whether vertebrates would, like infants, show a narrowing of their phonetic discrimination abilities given more exposure to a specific human language. One way to test this would be to test discrimination of a specific contrast (i.e., the vowels /e/ and /ɛ/) in both dogs who were exposed to a language that possesses that contrast (i.e., English) and dogs who were exposed to a language that does *not* possess the contrast (i.e., Spanish). If both sets of dogs succeed in the discrimination task, it would suggest that experience with a language does not narrow their perception of these phonological categories. If instead, only the English-hearing dogs succeed while the Spanish-hearing dogs fail, it would suggest that their phonological categories do change based on language exposure; we would expect that English-learning infants around 12 months old would succeed at this task as well, while Spanish-learning infants at 12 months would fail.

***Learning word forms.***

When infants build a lexicon, they need to learn word forms and also develop mappings from these words onto concepts and objects. Researchers have proposed that young infants first segment the speech they hear and develop a lexicon of word forms, which they can then map to referents (Hollich, 2006); however, young infants learn word forms even when there is not yet any meaning to map onto them. Infants also learn some word forms that are presented in isolation (Brent & Siskind, 2001); these learned words can then aid in segmentation of other words in fluent speech (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005). Studies have also shown that prior exposure to word forms facilitates the mapping of these word forms to objects (e.g., Estes, Evans, Alibali, & Saffran, 2007). By the time they reach eight months, infants have been exposed to thousands of word forms and will likely be able to recognize approximately a few hundred (Swingley, 2005). The learning of these word forms before connection with concepts plays a significant role in vocabulary development.

Few studies differentiate between the learning of word forms and word-object mapping in dogs. Mallikarjun, Shroads, and Newman (2019) found that dogs can recognize their own name and prefer to listen to it over another dog's name; this shows that dogs can learn word forms that are not necessarily mapped to specific objects or actions. This study found that there is no correlation between dogs' length of time with name and dogs' name recognition abilities; since the study required that dogs have their name for at least a year, this finding demonstrates that dogs need a year or less to successfully encode a word form. This, of course, does not significantly narrow the time frame required for a dog to learn a word form, given that infants can learn novel word forms after a few minutes of exposure to the novel words (i.e. Saffran, Aslin, & Newport, 1996; Thiessen, Hill, & Saffran, 2005). It is unclear what frequency and what type of input is necessary for dogs to learn word forms; future studies may test dogs in artificial-word-learning paradigms similar to Saffran, Aslin, and Newport (1996) to compare their abilities with those of infants.

***Word-object mapping.***

Word-object mapping involves the association of a word form with an object or action. While both infants and dogs can connect word forms to objects and actions, after their first year of life, infants' abilities quickly surpass those of dogs, as they can rapidly pair words and their referents after only brief exposure. Dogs, however, always require longer exposure time and more effort to successfully map words to their referents, regardless of age. While word-object mapping is not tested in the proposed studies, most studies of dogs' linguistic abilities involve word-object mapping, and it is relevant to their ability to encode word forms and recognize familiar words.

It is important to note that it is difficult in both infants and dogs to accurately measure vocabulary size for words that have been mapped to an object or concept; additionally, it can be difficult to assess what it means to be familiar with a particular word. Our lab's survey of pet dogs' vocabulary items indicates that their owners believe the dogs know approximately 5-15 words; in contrast, 10- to 13-month-old infants know between 38 and 100 words (Frank, Braginsky, Yurovsky, & Marchman, 2017). Some surveys of dogs' vocabulary have suggested instead that dogs know around 150 words (Coren, 2009); similarly, the MacArthur-Bates Communicative Inventory vocabulary norms suggest that 8-month-old infants recognize approximately 8 words, while other studies suggest that they can recognize hundreds of words (Swingley, 2005). The large differences in vocabulary size for both infants and dogs as suggested by these studies demonstrate the difficulty of properly assessing known words in these populations. However, it is clear that while dogs can learn words, it is a much more effortful process for them than for infants; and infants quickly outpace dogs with their word learning capabilities.

Early in infant development, word learning is slow and laborious, requiring numerous instances of the word form appearing with an object to successfully pair the two. Young infants can do word-object mapping with frequently heard items, like body parts (Tincoff & Jusczyk, 2012). They are capable of pairing these word forms to objects because they hear these word forms very often, which can facilitate their ability to map the word forms onto a concept. At 13-14 months, infants have been shown to link a new word to an object after only minimal exposure, also known as *fast mapping* (Werker, Cohen, Lloyd, Casasola, & Stager, 1998; Woodward & Hoyne, 1999; Woodward, Markman, & Fitzsimmons, 1994); however, these effects can be small and task-dependent (see Bloom, 2000, for a discussion). It is also not clear that the results of fast-mapping produce the same semantic richness that is achieved when a word is fully mapped into the semantic network (see Horst & Samuelson, 2008). These changes in infants' word learning abilities suggest that infants' word learning processes in the first two years improve over time, from initial laborious word-object mapping, to fast mapping. This period of time also coincides with the development of the infant phonetic system.

Domestic dogs, like infants, can map word forms to actions or objects. Word-object mappings and word-action mappings are often taught to dogs via multiple repetitions of the name in the objects' presence, playing with the object while the name is repeated, practice trials where the new object is placed amongst old objects and the dog is asked to retrieve the object by name (Griebel & Oller, 2012). Several exceptional dogs have learned between 200 and 1,000 discrete words for objects in this manner (Griebel & Oller, 2012; Kaminski et al., 2004; Pilley & Reid, 2011). It is important to note that most of these dogs take several years to acquire this vocabulary, whereas infants achieve a 200 word vocabulary by 15 to 17 months of age, and a 1,000 word vocabulary by 3 years (Frank et al., 2017; Shipley & McAfee, 2015). The dog Chaser often serves as a prime animal example of word learning, as she remembered the names of more than one thousand different objects and could retrieve them on command (Pilley & Reid, 2011). Chaser could also group these items by category, as well as pair specific actions with specific items (for example, putting the command *paw* or the command *nose* together with different familiar objects such that she could *paw piggie* or *nose ball*); this indicates that she could use words as referents for specific objects and actions (Pilley, 2013; Pilley & Reid, 2011). However, after nine years, Chaser had not expanded her vocabulary much beyond 1000 words, while children can easily attain much larger vocabularies as they get older. Chaser was a border collie, a working dog with exceptional drive. Another border collie, Rico, also had a large vocabulary of over 200 words (Kaminski et al., 2004). Since Chaser and Rico were initially the only two known dogs with large vocabularies, it was unknown whether the ability to learn words was exclusive to working breeds like border collies or whether other dogs also possessed the ability to learn many words. Griebel and Oller (2012) reproduced the Kaminski, Call, and Fischer study with a Yorkshire Terrier, Bailey. Yorkshire Terriers are traditionally lap dogs and do not have as much energy and innate desire to work as a Border Collie; however, Bailey demonstrated that she knew the names of 200 different toys and was able to retrieve them on command.

The majority of word-learning studies in dogs focus on exceptional, well-trained dogs with large vocabularies. In the proposed studies, I will test pet dogs and a small subset of working dogs, such as service dogs, police dogs, and search-and-rescue dogs. These dogs, including the working dogs, all have vocabularies much smaller than that of Chaser, Rico, and Bailey. While there are no studies of the average pet dog's vocabulary, examining frequently-used training programs by the American Kennel Club can give a general idea of what the typical dog knows. American Kennel Club's puppy training program includes the commands *sit* and *down,* and the adult dog program teaches *come, stay,* and *heel.* Dogs also likely acquire some frequently heard words in their environment, as owners often report that their dogs know words related to food, walking, and other activities the dog enjoys. However, there are no studies that have examined this. One way to circumvent the difficulty of ensuring that multiple dogs are familiar with test words is by utilizing the dog's own name as a test item; studies have shown that dogs can recognize their own name, even when spoken by an unfamiliar voice (Mallikarjun et al., 2019).

While both dogs and infants initially have the ability to discriminate between speech sounds and learn word forms, infants' capabilities quickly surpass that of dogs. However, the linguistic abilities of infants in their first year and adult dogs as they relate to word form recognition and word representation are more similar. Dogs and infants both demonstrate categorical perception of speech sounds, and can recognize familiar word forms (of which they may know a similar number). As such, through the use of a comparative dog model, researchers can answer questions about whether certain mechanisms underlying speech perception are human-specific, and also examine the role of linguistic experience in the absence of a human linguistic system.

### Attention.

*Attention* refers to the selection of specific stimuli in the environment to focus on and further process. Given the vast amount of sensory input in our environments, it is impossible to process all the information available to us at once. As a result, some information is privileged over other information, either by choice of the organism in the environment or due to the saliency of the information (Kagan & Lewis, 1965).

Attentional processes play an important role in speech perception. Listeners must utilize attentional abilities to selectively focus on the target stimulus while ignoring any existing distractors, and maintain that attention on the target stimulus for a period of time (Gomes, Molholm, Cristodoulou, Ritter, & Cowan, 2000). In order to accurately compare dogs' and infants' behavior in speech perception tasks, it is necessary to assess their similarities and differences in attention. In this section, I review infants' and dogs' attentional processing. I then compare infants' and dogs' attentional capabilities and how their similarities and differences might be exploited to learn more about infant speech perception.

#### *Sustained attention.*

Sustained attention refers to the ability to maintain focus on a stimulus over time. It is necessary to compare infant and dog capacity for sustained attention to human speech to ensure that differences in looking time and preferences in studies are due to the underlying process in question, and not overall differences in sustained attention.

Two studies have explored the development of sustained auditory attention in older children and have found conflicting results. One study examined 7- to 12-year-old children’s performance in an auditory vigilance task and found that the decrease in attention over the course of the task did not vary as a function of the children’s age (Gale & Lynn, 1972). Another study found that eight-year-old children’s sustained attention during an auditory vigilance task deteriorated more quickly than older children between 10-15 years of age (Swanson, 1983). Studies of auditory sustained attention in children who are pre-school-aged and younger are necessary to better determine the time-course of development.

While there are no current studies examining sustained auditory attention in infants, there are many studies that indirectly examine sustained attention, as they measure the amount of time infants listen to different auditory sources (e.g., Barker & Newman, 2004; Newman, 2005). However, the purpose of these studies is usually to determine infants’ preferences rather than their overall looking time, and the studies do not include information about the number of times infants disengaged from and reengaged with the auditory source. In dogs, most sustained attention tasks explore dogs’ ability to attend to humans while they help with service and therapy tasks (e.g. Alterisio et al., 2019; Cavalli, Carballo, Dzik, Underwood, & Bentosela, 2018; Mongillo, Bono, Regolin, & Marinelli, 2010). These tasks show that with training, the length of time dogs can sustain attention on their owner increases (Mongillo et al., 2010). In Mallikarjun, Shroads, and Newman (2019), we observed that dogs who are trained for specific jobs or tasks, like police work or service, show a larger listening bias for their own name in comparison to a foil name; this may suggest that they are better at identifying and sustaining attention on relevant stimuli. As such, it is important to balance the number of dogs that have extensive training in obedience, service, or search-and-rescue in comparative studies as these highly-trained dogs may have better attentional abilities than pet dogs.

To directly compare dog and infant sustained attention for auditory stimuli, overall looking times to stimuli over time were examined in dog data from Mallikarjun et al. (2019) and infant data from Newman (2009). This analysis showed that overall, dogs attend for shorter periods of time to auditory stimuli in comparison to infants. However, dogs and infants show a comparable drop-off in looking time across blocks of stimuli, which demonstrates that their attention for auditory stimuli over time is similar (see Figure 3; Mallikarjun et al., 2019; Newman, 2009). Further studies should evaluate dog and infant auditory sustained attention using a more specific task that directly assesses this skill, like the auditory vigilance task used with older children.

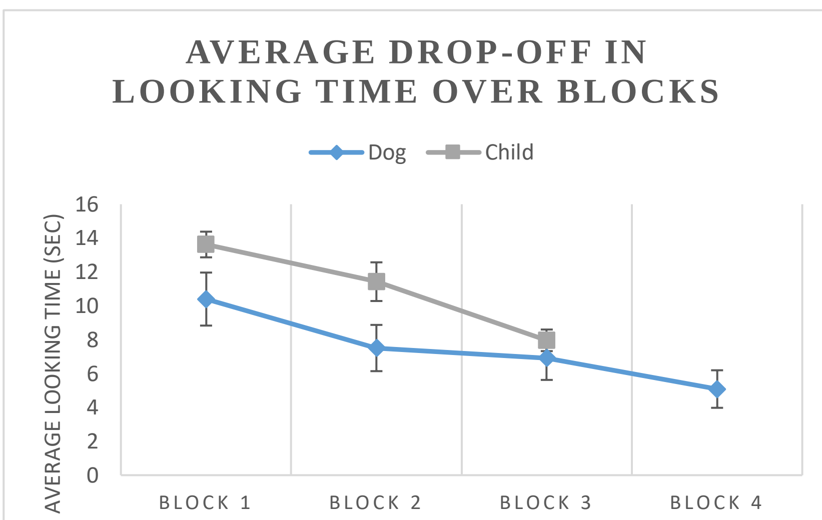


Figure 3: Comparison between dogs' and 8-month-olds' looking times for speech stimuli over time as measured by different experimental blocks of stimuli, with data from Newman (2009) and Mallikarjun et al. (2019).

***Selective attention.***

In addition to overall looking time to auditory stimuli, it is important to assess dog and infant preferences for certain types of stimuli over others, as well as their ability to attend to target stimuli in the presence of simultaneous background streams. *Selective attention* refers to the process of focusing on one stimulus while ignoring or attenuating other distracting stimuli. This can be driven either by conscious choice, or automatically due to the salience of certain stimuli. In order to demonstrate selective attention, a participant must preferentially attend to one stimulus over another stimulus. Selective attention is typically demonstrated in tasks in which the participant is verbally directed to select and attend to a specific stimulus to the exclusion of the other stimulus. As such, selective attention can be difficult to study when the participant will not respond to verbal commands. Both infants and dogs cannot follow verbal direction in a study; as a result, much of what is known about selective attention in infants and dogs is derived from studies that were primarily investigating other cognitive processes.

Early in development, infants do show preferences for certain sounds over other sounds. For example, infants generally prefer to listen to infant-directed speech over adult-directed speech (Fernald, 1985; Pegg, Werker, & McLeod, 1992), their own name over other infants' names (Mandel et al., 1995), and utterances with native-language prosody rather than foreign-language prosody (Mehler et al., 1988). Infants' preferences depend on factors such as stimulus complexity and familiarization time (DePaolis, Keren-Portnoy, & Vihman, 2016) as well as age, where older infants tend to prefer novel stimuli to familiar stimuli as the familiar stimuli becomes increasingly boring (Hunter & Ames, 1988). Dogs also listen longer to some sounds than others; in recent studies, dogs listen longer to dog-directed speech in comparison to adult-directed speech (Ben-Aderet et al., 2017) and their own name in comparison to another dog's name (Mallikarjun et al., 2019). While overall listening times to stimuli itself is only a part of selective attention, knowledge of infants' and dogs' listening times for certain stimuli can be used in selective attention studies to demonstrate their ability to focus on one preferred auditory stream while ignoring another stream.

The ability to selectively listen to one stream in the presence of another stream is a slow-developing skill in infants. Even at 8 years of age, children's selective attention is not yet adult-like (Coch, Sanders, & Neville, 2005); as such, it is expected that infants' selective attention is also not adult-like. Infants' selective attention tends to be less selective in general than adults' selective attention. For example, one study compared infant and adult selective attention for expected tone frequencies presented in background noise (Bargones & Werner, 1994). Adults selectively listen for the expected tone frequency in background noise. In contrast, infants do not selectively monitor for this expected frequency, instead maintaining a wide band of attention across frequencies. This suggests that infants are less selective in their auditory attention; however, it could also imply that infants do not have the same kinds of strategies adults have when listening in background noise. Studies have shown that infants are capable of processing information from one voice, or one auditory stream, in the presence of other simultaneously occurring streams. Infants older than 7.5 months are generally capable of paying attention to one voice in background multitalker babble when the target voice is more intense than the babble (Newman & Jusczyk, 1996; see Figure 2). In this study, infants were first familiarized with repetitions of two target words while a distractor passage played. The distractor passage was 10 dB less intense than the target words. During the test phase, infants heard passages containing the target words as well as passages containing unfamiliarized words. If the infants successfully separated the target words from the distractor passage during familiarization, they should have listened longer at the test phase to passages containing the target words than passages containing unfamiliarized words. This was the pattern that infants demonstrated, showing that they could successfully separate target words from simultaneously presented auditory stimuli. Likewise, at 13 months, infants recognize their name in multitalker background noise if the name is more intense than the background (Newman, 2005). These studies suggest that infants are capable of selectively attending to one voice, or one auditory stream, in the presence of another.

Dogs also can selectively attend to a voice in the presence of other auditory streams. In Mallikarjun, Shroads, and Newman (2019), dogs preferred to listen to their own name in comparison to another dog's name, both in quiet and in nine-talker background noise at the same intensity as their name. This study is very similar to Newman (2005). Comparing the dogs in Mallikarjun et al. (2019) and the infants in Newman (2005) can help determine what stimuli are interesting to dogs and infants and what they attend to longest. Figure 4, below, displays graphs from both Newman (2005) and Mallikarjun et al. (2019). The first two columns of the graph on the left show infants' looking times for their own name and a stress-matched foil in 5 dB

SNR 9-talker background noise. The graph on the right shows the looking times for dogs' own name and a stress-matched foil in the same background noise. To determine the similarity of effect size for dogs and infants, I examined dogs' and 13-month-old infants' average listening times for their own name in 5 dB SNR 9-talker background noise. These specific studies were chosen because of their similarity to the studies in Chapter 1 of this dissertation. The individual effect sizes as measured by Cohen's $d$ ($d$ = .37 for dogs, from Mallikarjun et al., 2019; $d$ = .35 for infants, based on Newman, 2005) suggest that the degree to which dogs and infants prefer and sustain attention to their name in comparison to the foil is similar.

It is worth noting that in these studies, infants and dogs may be auditorily segregating the streams and then attending to both of them simultaneously, rather than attenuating the background stream. There are no current studies that specifically address this. To examine this issue, researchers could examine infant and dog listening times for a stream containing their own name that co-occurs with either a non-meaningful background stream, which contains no familiar words, or a meaningful background stream, that contains more interesting known words. If infants listen longer overall to stimuli with the meaningful background stream, it would indicate that the participants are allocating some attention to the background noise.

Overall, the developmental timeline of the attention system in infants is long, and infants do not reach maturity in most attention-related skills in the first year of life. Infants' ability to selectively focus on particular sources of information and to sustain their attention for periods of time is still immature. While dogs' attentional abilities have not been studied to the same degree as infants, our studies have demonstrated that dogs' selective and sustained attention in speech perception tasks is relatively similar to that of infants.

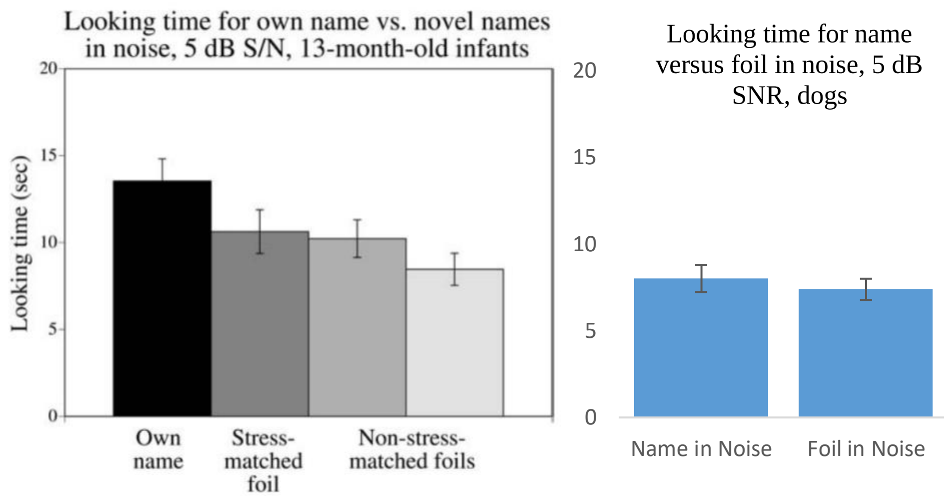


Figure 4: Left image, from Newman (2005) shows 13-month-old infants' average looking times towards their name and foils in 5 dB SNR 9-talker background noise. Right image is data from Mallikarjun (2019), rescaled for comparison with the Newman (2005) graph. The graph on the right shows dogs' average looking times to their name in +5 dB SNR 9-talker background noise, and a foil in the same level of noise. The background noise used in both studies was the same; however, the infant experiment did not feature trials in quiet, because it was already known that infants would respond to their name in quiet. It was not known when the study from Mallikarjun et al. (2019) was conducted whether dogs would respond to their name in quiet in an experimental setting. It is clear that dogs do not attend as long as infants for any trial type, but they do show a similar pattern of listening longer to their own name in noise than the foil in noise.

**Auditory system.**

The auditory system detects, segregates, and processes sounds present in the environment. In mammals, it consists of sensory structures, including the outer, middle, and inner ear, as well as neuronal structures that take sensory information and utilize it to understand features of the input like rhythm, number of auditory objects, and sound origin location. In this section, I will compare infants' auditory processing skills to dogs' auditory processing skills. Dogs' auditory abilities are more similar to that of adult humans than infants. Infants are still developing their auditory processing skills at 13 months of age, which is the target age for the proposed studies, below. Given the differences in the dog and infant auditory system as well as their similarities in attentional skills, comparison of dogs' and infants' performance in a similar speech perception task can identify whether the task relies primarily on auditory processing abilities. If dogs outperform infants in a speech perception task, the task may rely more heavily on auditory processing as compared to attentional processing.

***Hearing range.***

Adult humans and dogs have an overlapping hearing range and similar auditory thresholds in the frequency range that encompasses speech. Dogs have a hearing range from 67 Hz 45 kHz, which overlaps with humans' hearing range from 20 Hz to 20 kHz. However, dogs can hear sounds at lower intensities than infants can, and can also hear sounds at higher frequencies. Both dogs and adults have similar thresholds between 500 Hz to 10 kHz (Strain, 2011): see Figure 5, below, for a graph of dog, adult human, and infant thresholds. However, dogs are able to better detect sounds at frequencies between 10-20 kHz; they can perceive stimuli that are an average of 14 dB SPL less intense than normal-hearing adult humans at these frequencies (Lipman & Grassi, 1942). Additionally, given dogs' ability to detect high frequency sounds, it is necessary to check auditory stimuli for the presence of high frequency artifacts (see Brandt & Bitzer, 2014), and electronic equipment for any interference. This extraneous noise would present a potential distraction for the dog during a study.

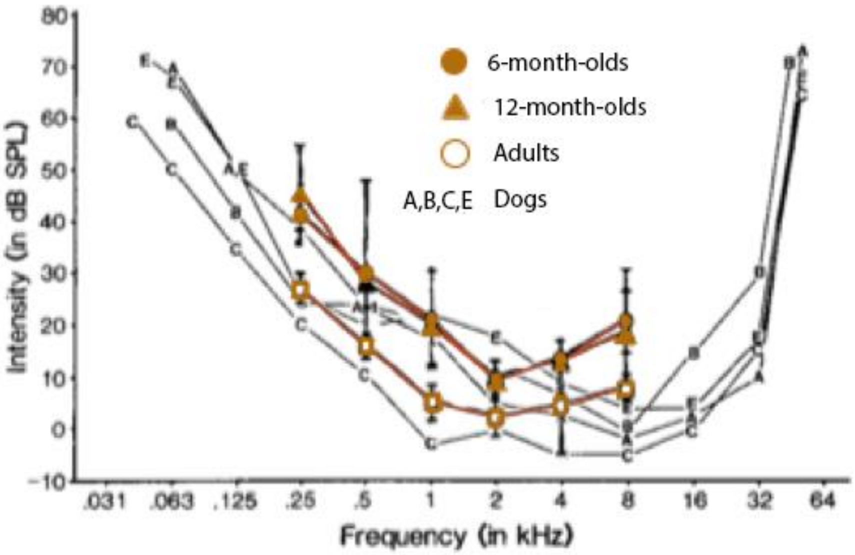


Figure 5: An overlay of part of the threshold graph from Olsho, Koch, Carter, Halpin, & Spetner, (1988) depicting pure-tone auditory thresholds in 6- and 12-month-old infants as well as adults, on top of the Heffner (1983) dog audiograms. Human audiograms are in orange, while dog audiograms are in black. Between .25 kHz and 4 kHz, infants' thresholds are generally higher than adults' thresholds. In the same range, dogs' thresholds vary, but are relatively similar to the human thresholds. Around 8 kHz, the dogs have better thresholds than the humans.

### ***Frequency discrimination.***

The ability to identify that two pure tones played sequentially are different from one another is known as *frequency discrimination.* The ability to discriminate frequencies is necessary to identify different streams of sound. When sounds become less intense, they are more difficult to hear, and as a result, become more difficult to differentiate. As such, sounds need to vary more in frequency to detect a change if the sounds are less intense (Nozza, 2005). Given that young infants have poorer thresholds for sound than adults and need sounds to be more intense to hear them, they should also have poorer frequency discrimination than adults.

Infants do not possess mature frequency discrimination abilities. Olsho and colleagues demonstrated that infants could reliably detect frequency changes when the degree of change was at least 2%, while adults could detect differences when the degree of change was only 1% (Olsho, Schoon, Sakai, Turpin, & Sperduto, 1982). This study, however, tested infants at a relatively difficult intensity level. Olsho, Koch, and Halpin (1987) expanded upon this study by testing infants and adults using dB SL, which means the intensity level presented to each participant was relative to the participant's sensory experience of the sound. This accounts for the fact that infants' thresholds are worse than adult thresholds for sound detection. In the frequency discrimination task, 6- and 12-month-old infants' performance was still worse than adults at 500 and 1000 Hz, but similar at 4000 Hz. An analysis of infant and adult psychometric functions for frequency discrimination shows that while infants' slopes are not as steep as the adults', and their performance asymptotes (reaches ceiling) at a lower point than the adults, the infant functions nevertheless behave similarly to the adult functions. It is important to note that these tasks inherently require attention; some percentage of infants' performance deficit (though not all) is likely due to differences in attention and not to differences in auditory function (Olsho, Koch, & Halpin, 1987; see Figure 6, below).

There are very few studies, and no studies done in the past 75 years, that directly explored dogs' ability to discriminate between different frequencies. One study used Pavlovian salivary conditioned response to test whether dogs could distinguish between a conditioned frequency at 637 vibrations per second and other frequencies. Dogs easily discriminated between frequencies that were close to one another in pitch, approximately 50-100 Hz apart (Anrep, 1920). Using a more contemporary conditioned response paradigm, such as a conditioned headturn or a go/no-go task, could better determine dogs' abilities to distinguish between close frequencies and allow for better comparison with human performance.

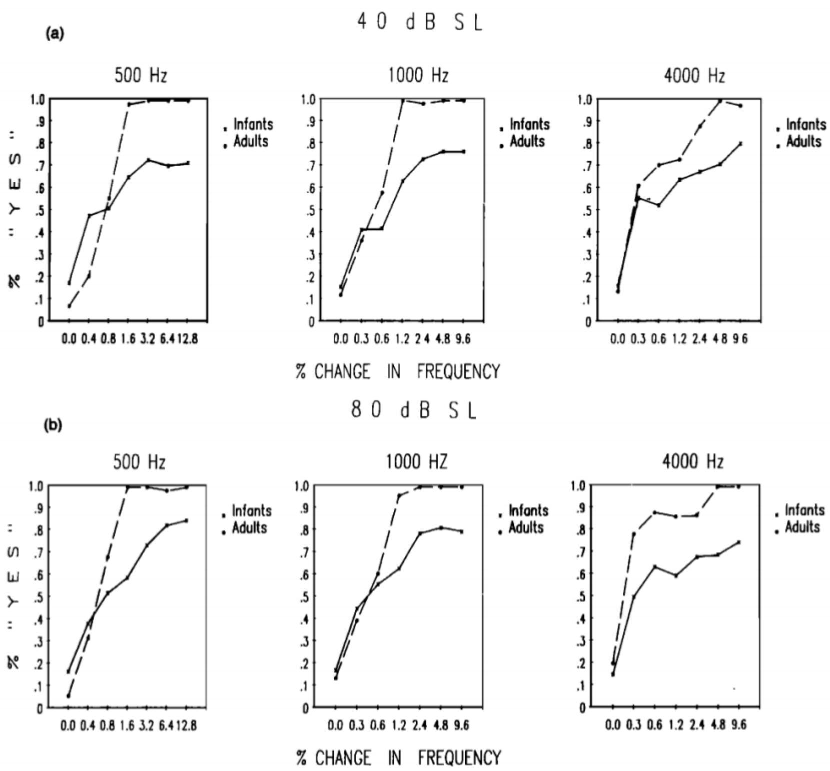


Figure 6: A graph from Olsho et al. (1987) showing psychometric functions for frequency discrimination (500 Hz, 1000 Hz, and 4000 Hz) at 40 dB SL and 80 SL. The infants' slopes are shallower than the adults and their performance reaches a ceiling at a lower point than the adults. However, the infant functions behave similarly to the adult functions. Olsho et al. suggest that part of the slope and asymptote differences are actually due to differences in attention between infants and adults.

***Sound localization.***

Another important auditory skill is localization, which is the ability to spatially orient to a sound source. Dogs' ability to localize a sound is similar to that of an adult human (R. S. Heffner & Heffner, 1992); infants, in contrast, do not demonstrate adult-like localization until after 18 months of age (Morrongiello, 1988). Sound localization is an especially important skill for dogs to locate prey and identify any threats in the environment. Shortly after puppies' ear canals open, they begin to localize single-source sounds, like a dog barking, at sixteen days old (Ashmead, Clifton, & Reese, 1986). Few studies explore auditory localization in the adult dog. One study confirmed the importance of the auditory cortex for sound localization by testing dogs with and without bilateral auditory cortex ablation on a sound localization task. Dogs heard a click and then had to approach the loudspeaker that produced the click out of a set of loudspeakers positioned around the room. Dogs with auditory cortex damage struggled at this task, while normal-hearing dogs achieved high levels of accuracy (H. E. Heffner, 1978).

Infants, like dogs,  will orient their heads towards single-source sounds shortly after birth, demonstrating a basic ability for directional discrimination (Clifton, Morrongiello, Kulig, & Dowd, 1981). However, this ability is still immature; newborns will reliably orient to human speech, but fail to orient to brief sounds (Clarkson, Clifton, Swain, & Perris, 1989). From 2-3 months, infants are transitioning from a subcortical, reflexive head-orienting response to a more adult-like cortical mechanism for localization (Muir & Clifton, 1985). In this transitional period, infants will fail to demonstrate any localization. Head-turning behavior towards a sound then reliably reemerges between 4-5 months (Muir, Clifton, & Clarkson, 1989). This U-shaped development is attributed to cortical maturation (Muir & Clifton, 1985).

In the first two years, infants dramatically improve in their precision of sound localization. They no longer just demonstrate directional discrimination; they also can more accurately localize sounds in the horizontal plane. At birth, infants can localize sound within 27 degrees (Morrongiello, Fenwick, Hillier, & Chance, 1994). By 18 months, this precision improves to 5 degrees (Morrongiello, 1988). This is still below adult performance between 1-2 degrees (Ashmead, Clifton, & Perris, 1987). While this is affected by physical factors, like the ability to accurately move the head, the largest improvement in precision occurs substantially after infants have developed reliable head control (Morrongiello & Rocca, 1987), which suggests that an improvement in hearing ability is responsible for infants' improvement in localization. As such, any comparative studies with dogs and infants involving more precise localization should utilize older infants that have more mature auditory localization abilities.

***Temporal resolution.***

Temporal resolution is a broad topic that relates to the detection of changes in time and perception of sounds that vary in time. Fine temporal resolution is important for language comprehension, as speech rapidly unfolds in time and contains many sounds, like consonants, that are of extremely short duration. Temporal resolution is also critical for *gap detection*, which is discussed below. Studies have suggested that children only reach adult performance in temporal resolution between 8-11 years of age, depending on the task. Dogs can use fine-grained temporal information to comprehend conspecific vocalizations (Siniscalchi, Lusito, Sasso, & Quaranta, 2012) and their auditory evoked responses demonstrate that they categorically differentiate between consonants (Adams et al., 1987; but see Chapter 5 for behavioral evidence that dogs do not detect consonant changes in their own name). However, infants are still developing their temporal resolution, which affects several aspects of speech perception, including their ability to differentiate between sounds, as well as their stream segregation abilities (Stuart, 2005).

One aspect of temporal resolution that is important for speech comprehension is a listener's ability to detect two temporally separated auditory events as separate instead of grouping them as a single event; this is known as *gap detection* (Williams & Perrott, 1972); this is crucial for distinguishing between stop consonants in words, as the presence or absence of an acoustic gap, as well as the length of the gap, can allow listeners to identify stop consonants in speech (Phillips, Taylor, Hall, Carr, & Mossop, 1997). Infants at 3, 6, and 12 months of age are much poorer at detecting the presence of a gap than adults. The 12-month-old infants, however, have a much higher variability in their performance, with some infants performing close to adult levels, and some performing more akin to 3- and 6- month-olds (Werner, Marean, Halpin, Spetner, & Gillenwater, 1992). It is difficult, however, to disentangle the effect of infants' developing auditory skills from their attentional abilities in this task. At 5 years old, children still demonstrate deficits in gap detection (Trehub, Schneider, & Henderson, 1995). At ten years of age, children can detect gaps in noise as short as 3.5 milliseconds (Phillips, Comeau, & Andrus, 2011) which is comparable to adult performance (Musiek et al., 2005).

While there is not much information about dogs' use of temporal resolution or gap detection in human speech comprehension, there are a few studies that examine dogs' use of temporal information when processing other meaningful sounds. Siniscalchi et al. (2012) indirectly examined whether temporal information was important in dogs' recognition of other dogs' vocalizations. Normal vocalizations and time-reversed vocalizations were played from two speakers oriented directly to the left and right of the dog participants (see Figure 7). Time-reversed vocalizations have the same spectral envelope, but different temporal features than normal vocalizations. Dogs were successfully able to discriminate between normal and time-reversed dog barks; therefore, temporal information must play some role in identifying communicative barks.

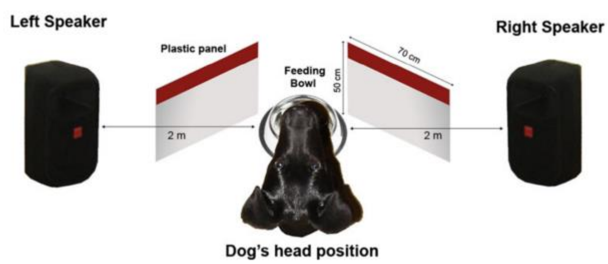


Figure 7: A depiction from Siniscalchi et al. (2012) of the testing setup in their study of the importance of temporal cues in conspecific vocalization.

In conclusion, in the areas that have been examined to date, the dog auditory system is more similar to the human adult auditory system than the infant auditory system, which is still developing. However, there are many aspects of dog auditory processing that have not yet been examined. While dogs cannot function specifically as a model of the infant auditory system, dog and infant auditory processing differences can be used to help assess whether certain speech perception tasks rely primarily on the auditory system, or whether they are more linguistic or attentional in nature.

### Caution in interpretation: Structural differences in dog and infant brain anatomy.

While there are identifiable behavioral analogies between certain underlying systems in dogs and infants, there is little research on the correspondence between human and dog brain anatomy and circuitry. Given that dogs and humans do not share a recent common ancestor, like humans and great apes, it is unclear whether dogs' analogous speech perception abilities stem from similar underlying neural structures developed from a common ancestor to dogs and humans, or whether these similar abilities emerged independently in dogs and humans from different structures. For example, one fMRI study examined whether dogs, like humans and great apes, have specific areas in the brain for processing conspecific vocalizations. Dogs do have a specific region for this purpose, and the location of this region overlaps with that of humans (Andics, Gácsi, Faragó, Kis, & Miklósi, 2014). However, while most of the human auditory cortex responded maximally for human-produced sounds, the dog auditory cortex favored environmental sounds over dog vocalizations. It is likely that humans have more specialized auditory processing regions for conspecific vocalizations than dogs do. In processing human speech, dogs show a hemispheric bias for processing lexical meaning, like humans; however, human processing for lexical meaning is left-hemisphere biased, while dogs' processing for lexical meaning is right-hemisphere biased. Dogs also possess a specific area for processing of intonation cues, but this was localized in a different area than that of human intonation processing.

These differences suggest that while the dog can be a useful behavioral model for speech perception, caution should be taken in interpreting dog and human results to suggest a particular course of evolution for these speech perception processes.

**Conclusions about generality.**

This section assessed the use of dogs as a model for infant speech perception using the model selection criterion of *generality*, or the extent to which specific research findings from dogs can apply to humans. First, dogs can be tested using methods similar to classic infant methods, which allows for better comparison between dog and infant results. Additionally, an examination of the linguistic, attention, and auditory systems in dogs and infants shows that infants have more linguistic capabilities than dogs given their human linguistic system, but dogs and infants have relatively similar attention for human speech. Conversely, dogs have a more sensitive auditory system than infants, who are still developing their auditory abilities. These similarities and differences can be utilized to design comparative studies that answer broad questions about the role of these systems in specific speech perception tasks. However, caution must be taken to ensure that generality is not confused with evolutionary similarity; currently, research demonstrates some structural differences between the dog and human brain, and further research is necessary to better establish the extent of the structural differences.

### ***Overall discussion***

Cross-species comparisons for speech perception are useful in shedding new light on the relative influences of linguistic experience and infants' various developing systems. In particular, investigating word recognition in a non-human species, like the domestic dog, that does not acquire language in the same way young children do, may help us to disentangle the contributions of auditory, attentional, and linguistic processing. The domestic dog is a particularly good developmental model of speech perception for several reasons. Dogs naturally have the ability to perceive speech, and they are extremely accessible as a model. Dogs' natural exposure to speech lends itself to comparative studies examining the role of speech exposure in the absence of a human linguistic system. Additionally, the similarities and differences in the underlying mechanisms responsible for speech perception in dogs and infants allows for a better understanding of the contribution of mechanisms underlying speech perception in infants. These similarities and differences allow for comparative studies that can tease apart underlying speech perception processes that would otherwise be impossible to differentiate by testing only infants and adults.

# Chapter 2: The Cocktail Party Effect in the Domestic Dog (Canis familiaris)

## *Overview*

Noise is ubiquitous in modern society: the sounds of airplanes, road traffic, and crowds can be found in most urban, public settings. A great deal of work has examined how adults cope with such environments, and more specifically their ability to understand speech in noisy settings. Yet adults are not the only ones facing this challenge; so too are both young children and our canine companions. How do dogs contend with noise when given commands from their owner, and what can this tell us about infant language comprehension in noise?

Dogs are an interesting population to study for several reasons. Dogs have co-evolved alongside humans to pay attention to human behavior. Dogs, like infants, pay attention to gaze, pointing gestures, and facial expressions, which all help dogs connect and communicate with humans (Albuquerque et al., 2016; Soproni, Miklósi, Topál, & Csányi, 2001). Their attentiveness extends not only to human behavior, but also human vocalizations. Dogs have brain regions specifically tuned to human vocal productions (Andics et al., 2014), as well as temporal area activation for human faces (Cuaya et al., 2016), and they use this information to determine emotional valence and meaning behind human language (Albuquerque, Guo, Wilkinson, Resende, & Mills, 2018, for emotion; Andics et al., 2016, for words). They are not only sensitive to humans' communicative behaviors, but make communicative bids of their own, making eye contact with humans to demand attention and communicate their needs (Merola, Prato-Previde, & Marshall-Pescini, 2012). Their direct ancestor, the wolf, does not do this, indicating that the domestication process and interactions with people have brought about this human-like behavior. Dogs' ability to recognize and respond to human communicative behaviors allows them to inhabit a number of roles in society, from companion animals in our homes to working as seeing-eye dogs, police dogs, search and rescue dogs, and more. Understanding dogs' ability to respond to human speech in difficult listening environments is important information for dog trainers, particularly for those who train service and working dogs, who must perform tasks in a variety of distracting environments and listening conditions. Dogs' social behaviors and attention to human communicative vocalizations and gestures also makes them ideal for use in comparisons with human infants and children.

Cross-species comparisons for word recognition in noise are useful in shedding new light on the relative influences of linguistic experience and infants' various developing systems. In particular, investigating word recognition in a non-human species that does not acquire language in the same way young children do may help us to disentangle the contributions of auditory processing and attentional systems from linguistic processing. Despite a large body of research documenting infants' and children's difficulty listening in noise, it remains uncertain what factors contribute most to individual differences in performance on speech-recognition-in-noise tasks. While immaturity in the auditory processing system could explain infants' poorer performance at listening-in-noise tasks, infants' basic auditory abilities are already adult-like by six months of age (for a review, see Werner, 2007). Their deficits could alternatively be explained by lack of cognitive maturity and relatively small linguistic and lexical knowledge, but it is difficult to tease apart these factors from auditory causes or from one another (Erickson & Newman, 2017). Using an animal model to examine speech perception in noise can aid in distinguishing linguistic and auditory factors, as animals do not have complex linguistic systems like humans and would be most affected by auditory, cognitive and attentional issues in speech perception.

Dogs are particularly well-suited for comparison with young children on speech-in-noise tasks. Dogs have the ability to quickly assign a label to a novel object and retain that connection in memory, as do young children (Kaminski et al., 2004). Work with individual dogs has suggested that some may acquire vocabularies that are similar in size to those of young children (Pilley & Reid, 2011). Dogs have evolved alongside humans to be particularly attentive to human behaviors and are highly socially motivated, characteristics that are useful in adapting existing research tasks. Several classic paradigms originally designed for young children have been utilized with dogs with minimal modifications (particularly tasks designed for preverbal children; see Fugazza & Miklósi, 2014). For example, one study looked at dogs' numerical understanding using the same preferential-looking technique and study design as an earlier study that examined infant numerical understanding. (West & Young, 2002, for dogs; Wynn, 1992, for infants). Another study that examined dogs' ability to recognize familiar human faces, dogs, and objects used a preferential looking paradigm in which the dogs were shown two images on a large television screen, which is similar to the design of many infant studies (Racca et al., 2010, for dogs; Rhodes, Geddes, Jeffery, Dziurawiec, & Clark, 2002, for face stimuli shown to infants).

The current work examines canine companion performance at understanding a spoken word in the presence of noise, using a very similar paradigm used to test infants' abilities. The ability to understand speech in the presence of noise is critical for both species. For dogs, this is most apparent when considering service dogs, who must face a number of different noisy environments with their handler. In cities, they will hear traffic, machinery, and constant low-level noise from pedestrians; it is also likely the case for pets, whose owners may call to them from a distance. Police dogs must also contend with gunfire, sirens, and loud voices. These noises can all compete for attention with the actual commands and tasks a service dog must perform, and if the dog does not pay attention properly, the dog can potentially endanger the handler. Anecdotally, these dogs perform very well in these situations and correctly complete their tasks when given commands from their handler. In this set of studies, we aim to quantify the level of background noise at which it becomes difficult for service dogs and pet dogs to pay attention to an important, salient word. We test dogs raised in a home environment, for whom attending to human speech is a natural behavior, as compared to dogs raised in a more impoverished lab setting (see Fugazza & Miklósi, 2014, for more on this point).

In addition to exploring how well dogs can understand speech in these environments, the current study also serves as a useful comparison to young humans. Infants and young children are notably poorer at speech recognition and language processing in the context of background noise compared to adults. Infants have poorer auditory thresholds for speech than adults, meaning that they need speech to be louder than adults would typically need before they can detect it (Trehub, Bull, &

Schneider, 1981). Greater speech intensity is also needed for infants to distinguish between speech sounds embedded in noise (Nozza, Rossman, Bond, & Miller, 1990, for noise; Nozza, Rossman, & Bond, 1991, for quiet). These limitations also occur when speech is in the presence of other environmental sounds (Polka, Rvachew, & Molnar, 2008) or background speech (Newman, 2005, 2009; Newman & Jusczyk, 1996). Infants between 5-8 months of age generally need the target speech to be louder than the background speech in order to comprehend it (Newman & Jusczyk, 1996).

It remains unclear whether the source of such difficulties is purely the result of poor auditory and attention skills, or might also be affected by having a limited language system. While some have argued that attention is a critical factor (Erickson & Newman, 2017), other evidence supports the role of language experience. For example, bilinguals perform worse than monolinguals at hearing-in-noise tasks, even if they are highly proficient in their languages, with data indicating that this deficit in performance may be due to slightly reduced experience with the language as compared to monolinguals (Schmidtke, 2016). By comparing infant performance with that of dogs, we can gain a better understanding of the relative role of auditory and cognitive skills versus language-specific skills in infants' listening-in-noise difficulties.

### *Experiment 1: Mild noise versus quiet (+5dB SNR)*

In order to identify whether a listener can comprehend speech in the presence of noise, it is first critical to find a speech sound that the individual would comprehend in quiet. For this study, we utilized dogs' own names as the critical stimuli. These names were spoken by a novel talker, either in quiet or in noise, in a manner nearly identical to previous work with infants (Newman, 2005, 2009). Although dogs often have a great deal of experience hearing their name, they generally only hear it spoken by a relatively small number of people. Using a novel talker meant that the dog would need to generalize their knowledge of their name across different speakers, the manner in which the recording is done would not be identical to the way the dog normally hears its name. If dogs can recognize their own name when spoken by a novel talker, they should listen longer to this name than another dog's name when both names are presented in quiet.

If dogs succeed at this generalization task when presented in quiet, then by presenting these same names in the presence of noise, we can identify whether the noise is sufficiently distracting to limit their performance. Instead of using white noise or another artificial noise, we instead used a background of nine voices blended together. Multitalker babble such as this is a background noise that dogs may encounter in many situations when a crowd of people is present, like restaurant patios or in parks. We examined dogs' ability to separate and attend to target speech while there are multiple voices speaking in the background. By varying the difficulty level of the background noise, we can examine dogs' speech-in-noise abilities in conditions in which infants are successful or unsuccessful on this same task. To start with, we examined a relatively low level of noise, one that is akin to the ambient noise inside an urban home (McAlexander, Gershon, & Neitzel, 2015).

**Participants.**

Twenty dogs (6 male) participated in the study. In order to be included, dogs must have had their name for at least ten months prior to participating. We excluded any dogs that were taking psychiatric medication, and dogs whose owners noticed any signs of hearing loss. On average, participating dogs were 4.37 years old, and had been hearing their names for 3.97 years (i.e., the dogs had not been recently adopted such that they received a name change). Three of these dogs were bomb detection k9s, and one was a search-and-rescue dog; the remaining sixteen were all pet dogs. Three dogs had a one-syllable name, two had a three-syllable name, and fifteen had a two-syllable name. Of the three-syllable name dogs, one had an unstressed-unstressed-stressed pattern and one had an unstressed-stressed-unstressed pattern. All the two-syllable dog names had a trochaic stress pattern (stressed-unstressed).

To determine whether performance differed by breed, we also collected owner-report information on dog breed, and sorted the dogs into the seven AKC breed group categories based on their breed, or in the case of mixed-breed dogs, the most predominant breed. We had one dog in the herding group, one dog in the hound group, one dog in the non-sporting group, two dogs in the terrier group, six dogs in the sporting group, five dogs in the toy group, and four dogs in the working group. Data from five additional dogs were excluded from the study: four for noncompliance (e.g. failing to orient to sounds, falling asleep), and one due to experimenter error. All dogs were tested in the presence of their owner, in order to reduce stress and ensure optimal performance (Fugazza & Miklósi, 2014).

**Test materials.**

Stimuli consisted of a target sound stream and a distractor sound stream. The distractor stream was the same as the multitalker babble used in the Newman (2005) study that examined infants' perception of their names in noise. For that study, nine women were recorded reading passages from books using a Shure SM51 microphone in a sound-attenuated room. These recordings were adjusted to be the same root-mean-square amplitude and then mixed together at equal ratios to create nine-voice multitalker babble. With this number of speakers, the babble converges to being a relatively constant intensity level over time. Moreover, it is impossible to make out individual words from this type of babble.

The target speech stream consisted of a name repeated 15 times: either the dog's own name or that of another dog. Prior to the study visit, each dog owner was asked the name or nickname that their dog was most commonly called. This name was recorded in advance of the appointment date and formed the target stream for the study stimuli. The names for each dog were recorded individually by a female native English speaker from eastern Pennsylvania. The speaker was recorded saying the dog's name in dog-directed speech with a variety of inflections and durations (Ben-Aderet et al., 2017). Each name was matched with a foil name. In order to prevent any bias caused by the speaker producing target names in a more lively manner, each foil was chosen from the existing set of recorded dog names, which were target names for other scheduled participants. The foil was matched to the target name in the number of syllables and stress pattern, and the names were chosen to be as phonetically dissimilar as possible from the original name in phonemes (e.g. *Henry* was matched with the foil name *Sasha*). A total of fifteen tokens were selected out of the original recording of target names, matched to the fifteen tokens of the foil name file as closely as possible for pitch, duration, intonation contour, emotionality, and vocal quality. Pauses between tokens of dog names were adjusted such that the target and foil files had the same overall duration of 22 seconds. There was an initial silence period for 0.5 seconds.

Intensity and amplitude of the target and foil name streams were measured and altered in order to match each other and in order to establish a set signal-to-noise ratio between the names and multi-talker babble. Name streams contained silence in between the name tokens, and although tokens were selected to have similar duration, the overall amount of silence in the target and foil streams was not necessarily identical. Therefore, in order to eliminate any influence of the silent periods on amplitude measurements, a copy was created of each name stream in which all the pauses between name tokens were removed. Average RMS amplitude was measured across this file, which contained only speech, and necessary amplitude changes were calculated and applied to the original stream containing pauses. In this way, the name streams could be amplified such that the *speech,* rather than the entire stream, matched in average amplitude. These streams constituted the “quiet” name stimuli.

In addition to the quiet name streams, each stream was mixed with a 22-second clip of the multitalker babble to create names-in-noise. Average RMS amplitude of the noise clips were set prior to mixing such that specific signal-to-noise ratios between the target speech (names) and babble were achieved. In Experiment 1, the noise was adjusted to be five decibels softer than the target speech (+5 dB SNR).

This is a level used in infant studies at which one-year-old infants are generally successful at name recognition (Newman, 2005).

**Apparatus.**

The experiment took place in a four-foot by six-foot three-sided test booth made of pegboard. In the front of the booth, there was a hole cut out for the lens of a video camera. Above the camera, a light was mounted in the center of the panel. The video camera recorded each session and allowed the coder to see the dog’s behavior inside the booth. The side walls each had a light mounted in the center and a speaker directly behind the light to play stimuli for the dog. A tan curtain hung from the ceiling to the top of the booth to ensure that the dog could not see over the booth. A Mac computer was used behind the front wall of the booth for coding. The experimenter used a button box to start trials and code the dog’s looking behavior.

**Procedure.**

The dogs sat on the owner’s lap or directly in front of them, depending on the dogs’ size and the owners’ opinion as to what would be most comfortable. The dogs either sat facing towards the camera (facing the front of the booth) or towards the owner (facing the back of the booth). In either case, the dog's attention was maintained (as much as possible) at a point equidistant from the two sides of the booth where the loudspeakers were located. As a result, the dog's natural inclination, upon hearing a sound through a loudspeaker, was to turn their head 90 degrees to face that sound source. There were two practice trials to familiarize the dogs with the procedure. In these trials, the dogs heard two different passages of classical music.

Their listening time was judged by the amount of time they spent looking at the sound source (the wall behind which the speaker was mounted).

The test phase began immediately after the practice trials. The dogs heard four types of stimuli: repetitions of their own name without background noise, a foil name, their name in the multitalker babble noise, and the foil in noise, four times each. The sixteen trials were presented in four, four-trial blocks, with one stimulus of each type per block, and order of trials within each block was randomized. Two experimenters ran the test-phase portion of the study, one to code the dog's looks (the coder), and one to produce auditory attention getters (the Attention experimenter). At the start of the test trials, the light on the front of the booth turned on, and the Attention experimenter rang a bell located behind the light. The combination of a light plus a bell served to attract the dog's attention.  Although work with infants typically uses only lights, pilot work suggested that neither the light nor the bell was a sufficient attention-getter for all dogs.  The light also served as the apparent source of the sound. Once the dog attended to the front, a light turned on in either the left or right side of the booth. The Attention experimenter rang a bell on that side. Once the dog attended to that side, the stimulus played from the speaker on that side. The stimulus played for a full twenty-two seconds or until the dog looked away for two consecutive seconds, whichever occurred first. Any time the dog spent looking away was subtracted from the dog's overall looking time. The coder used a button box to code the dog's looks towards and away from the sides. The coder wore Peltor aviation headphones playing masking music so she would not be able to hear the trials and have that influence her coding.

**Results.**

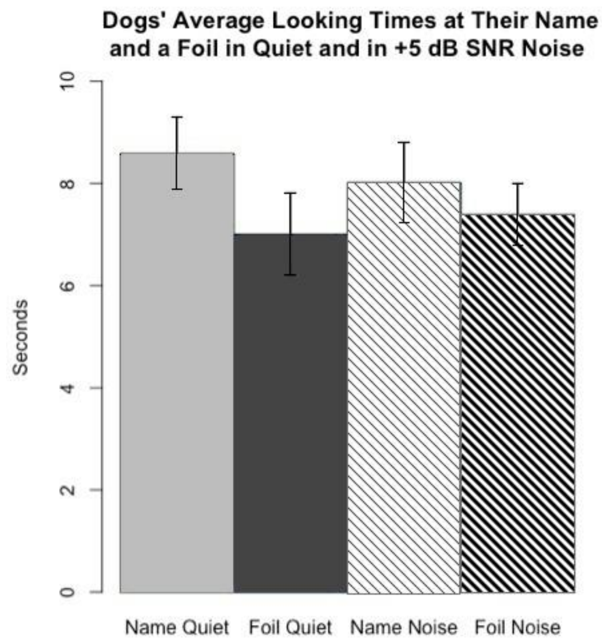


Figure 8: Dogs' performance in Experiment 1. Dogs listen significantly longer to their name than the foil.

Mean listening times were calculated for each of the four trial types (Name, Foil, Name in Noise, Foil in Noise) across the four blocks. Figure 8, above, displays dogs' mean listening times to the different trial types. A 2x2 analysis of variance (ANOVA) examined the effect of Noise Level (quiet versus +5 decibel signal-to-noise ratio) and Item (name versus foil).

We found an overall effect of Item, $F(1, 19) = 8.5$, $p < .001$, such that dogs listened longer to trials containing their name (8.3 seconds) than trials containing another dog's name (7.2 seconds). This suggests that dogs recognize their name, even when spoken by a novel talker. Thus, dogs are capable of generalizing known words across different talkers.

There was no overall effect of Noise Level ($F(1, 19) = .02$, $p > .05$); dogs listened just as long in quiet trials (7.8 seconds) as noise trials (7.71 seconds). Critically, there was no interaction ($F(1,19) = .59$, $p > .05$). That is, dogs' difference in looking times for their name over another name was the same in quiet as in noise. This pattern of results suggests not only that dogs recognize their own name, but also that the noise did not impact their ability to do so. Dogs apparently have little difficulty distinguishing their name from a foil name in either quiet or in the presence of this level of background noise.

This experiment showed that dogs are quite adept at generalizing language information across different talkers, and can thus successfully recognize their name as spoken by a novel voice. Moreover, since the names were matched for prosodic pattern, the dogs must be doing so based on the sounds or phonemes making up their name, rather than by the way the name was said (its emotional valence, or its pitch pattern). While there are clear anecdotal reports of dogs recognizing their name, this is the first time this has been shown experimentally in a task requiring generalization across talkers.

This experiment also showed that dogs also succeed at this task when in the presence of a quiet background babble. In the following study, we increased the level of the background distractor by 5 dB, resulting in a more difficult level of noise: 0 dB SNR. This particular level is useful for comparing canine performance with infant performance. Prior work has suggested that infants aged 13 months (but not aged 9 months) can succeed at this task at the +5 dB SNR tested in Experiment 1. However, infants at this age do not succeed with a 0 dB SNR. Thus, if dogs are successful, it would demonstrate that their ability to understand speech in noise is beyond that of a 1-year-old child.

### *Experiment 2: Target and background noise of equal amplitude (0 dB SNR*

Experiment 1 demonstrated that dogs were successful at recognizing their name when it was louder than the co-occurring background noise. The current experiment increased the level of the background noise by 5 dB. This resulted in the names and the noise being of equivalent amplitude.

#### Participants.

Twenty dogs (16 male) participated in this study. The dogs met the same requirements as in Experiment 1. They were an average of 5.3 years old. They had been hearing their names for 4.74 years on average. There were two dogs in the Herding group, one dog in the Non-sporting group, one dog in the Terrier group, five dogs in the Working group, and eleven dogs in the Sporting group. Fourteen of these dogs had jobs: four were therapy dogs, three were search-and-rescue dogs, one was a retired police dog, one was a service dog, and five were service-dogs-in-training.

#### Materials.

These were the same as in Experiment 1, except in the names-in-noise streams, the noise was adjusted to be equal in amplitude to the target speech (0 dB SNR).

#### Apparatus and procedure.

These were the same as in Experiment 1.

**Results.**

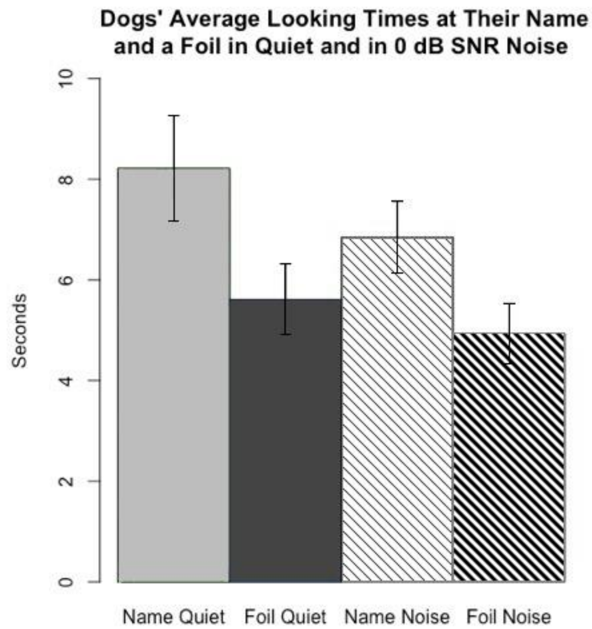


Figure 9: Dogs' performance in Experiment 2. Dogs listen significantly longer to their name than the foil.

The data were analyzed in the same manner as in Experiment 1. Mean listening times were calculated for each of the four trial types (Name, Foil, Name in Noise, Foil in Noise) across the four blocks. These listening times are shown in Figure 9, above. A 2x2 analysis of variance (ANOVA) examined the effect of Noise Level (quiet versus +0 decibel signal-to-noise ratio) and Item (name versus foil). We found an overall effect of Item, $F(1, 19) = 15.53$, $p < .001$, such that dogs listened longer to trials containing their name (7.52 seconds) than trials containing another dog's name (5.27 seconds).

There was no effect of Noise Level ($F(1, 19) = 2.28$, $p > .05$), as dogs listened just as long to items in quiet (6.91 seconds) as they did items in noise (5.89 seconds).

There was also no interaction between Noise Level and Item ($F(1, 19) = .25, p > .05$); this suggests that dogs continued to prefer their name to the foil despite the noise.

### *Experiment 3: Background noise louder than target (-5 dB)*

Experiment 1 demonstrated that dogs were successful at recognizing their name when it was louder than the co-occurring background noise. Experiment 2 showed that dogs were likewise successful at name recognition when their name and the background noise are of equal intensity. One-year-old infants do not succeed when the target is as loud as the background noise; since dogs succeed at this level, their ability to recognize their name in noise surpasses that of an infant. The current experiment increased the level of the background noise by 5 dB, resulting in the noise being louder than the target name. This will help determine at what point dogs fail to perceive their name in noise.

#### **Participants.**

Twenty-two dogs (11 male) participated in the study. They were an average of 5.3 years old, and had been hearing their names for an average of 4.74 years. Data from six dogs were dropped from this study. Two did not have their name long enough, and four were uncomfortable in the booth and the experiment had to be discontinued. Five of these dogs were service-dogs-in-training, and one was a therapy dog. Five dogs had a one-syllable name and the remaining 17 dogs had a two-syllable name. All the two-syllable dog names had a trochaic stress pattern (stressed-unstressed).

Two dogs in the hound group, six dogs in the non-sporting group, three dogs in the terrier group, eight dogs in the sporting group, and three dogs in the working group participated in this study. Data from five additional dogs were excluded from the study. Three dogs were excluded for noncompliance (e.g. failing to orient to sounds, falling asleep), one was excluded because the dog was too young, and one was excluded due to experimenter error.

**Materials.**

These were the same as in Experiment 1 and 2, except in the names-in-noise streams, the noise was adjusted to be 5 decibels louder than the target speech (-5 dB SNR).

**Apparatus and procedure.**

These were the same as in Experiment 1 and 2.

**Results.**

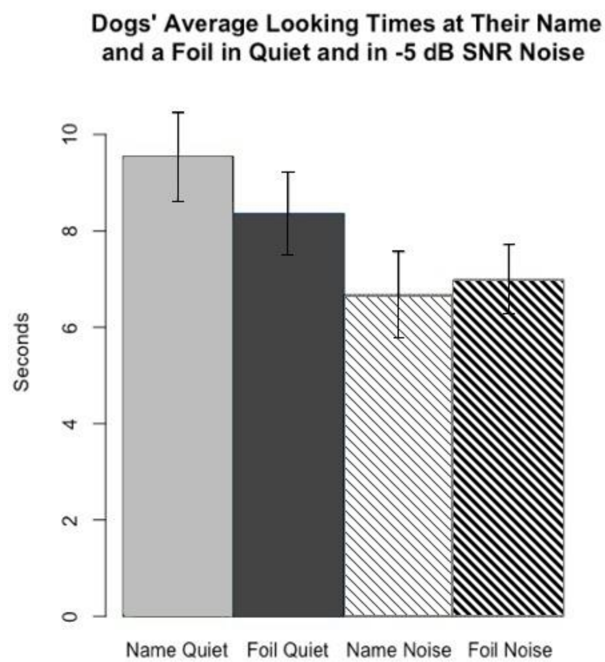

Figure 10: Dogs' performance in Experiment 3. Dogs listen significantly longer to the quiet trials than trials in noise.

The data were analyzed in the same manner as Experiment 1 and 2. Mean listening times were calculated for each of the four trial types (Name, Foil, Name in Noise, Foil in Noise) across the four blocks. The listening times are shown in Figure 10, above. A 2x2 analysis of covariance (ANOVA) examined the effect of Noise Level (quiet versus -5 dB signal-to-noise ratio) and Item (name versus foil). We found no significant effect of Item ($F(1,21) = .63$, $p > .05$). There was a significant effect of Noise Level, $F(1\ 21) = 6.199$, $p < .05$, such that dogs prefer to listen to the quiet items (8.9 seconds) more than the items in noise (6.8 seconds). However, there was no interaction between Item and Noise ($F(1,21) = 1.088$, $p > .05$).

Unlike in the prior two studies, the dogs here did not prefer their name to the foil name when the noise was present. Figure 11, below, shows differences in looking time to the name and foil in quiet and at all three noise levels. Dogs' lack of preference for their name over the foil in this condition might suggest that the level of noise presented here posed too much difficulty for the dogs. But surprisingly, the dogs also did not show an interaction between item and noise, implying that they also did not prefer their name to the foil name even in quiet. That is, the presence of the more difficult noise on some trials not only prevented the dogs from succeeding on those particular trials; it also prevented the dogs from succeeding at all. Why might this have occurred? One possibility is that the difficulty of the task led dogs to "give up" doing the experiment. Yet the dogs did not listen to all items equivalently – they preferred listening to both names in the quiet condition over those in the noise conditions. Perhaps this more intense noise was confusing or irritating to them, and also led them to stop attending to the detailed sound patterns within the name. Or, perhaps the loud noise caused them to attend to the background (the presence or absence of noise) rather than the target.

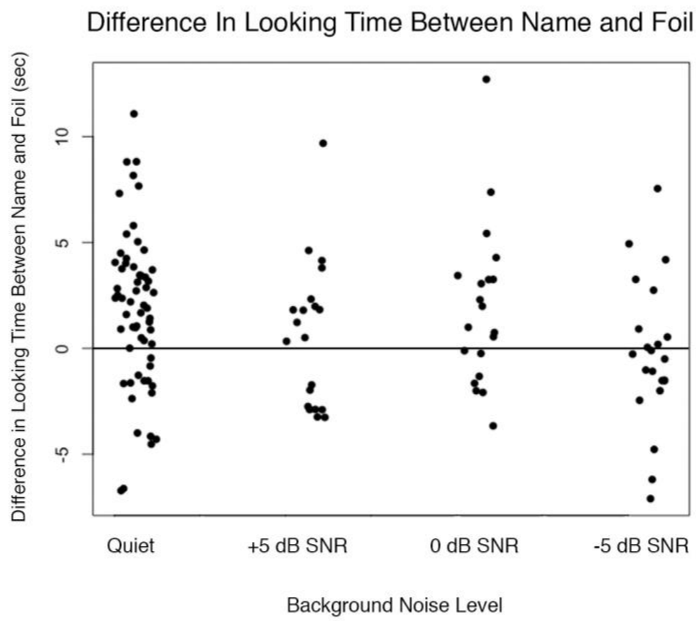


Figure 11: A graph of individual differences in looking time between the name and foil in quiet, +5 dB SNR, 0 dB SNR, and -5 dB SNR. The distributions of individual performance in the three noise conditions appear generally similar, but with lower performance in -5 dB SNR. There is no indication of greater variability in performance among dogs in the -5 dB SNR, as might be suspected if some dogs were succeeding at the task and others not. While 5 dogs shows scores that appear to be above chance performance, 3 dogs showed an equal performance below chance, suggesting this may have just been the result of random variability.

While we cannot be certain why the dogs failed in the present task, the results here are clearly quite different than those in the prior experiments. The level of noise presented here, -5 dB SNR, appears to be at a level sufficient to interfere with dogs' recognition of their name. Presumably, then, this level of noise would also pose problems for comprehending other speech sounds or commands.

### ***Breed-specific results***

Anecdotally, people have noticed that different breeds seem to have specific personality traits that lead them to respond to human speech differently. For example, one study showed that dogs who were bred for working purposes (Siberian Husky, German Shepherd) demonstrated more attentiveness to human communicative gestures than non-working breeds (Toy Poodle, Basenji) (Wobber et al., 2009). However, there have been studies of dogs' understanding of human gestural communication that show no breed differences (Dorey, Udell, & Wynne, 2009, no differences between the American Kennel Club groups; Mckinley & Sambrook, 2000, no difference between gundogs and non-gundogs). We examined whether performance differed by breed group. We also examined whether performance differed for dogs who were trained as working dogs vs. those that were pet dogs; other studies have suggested, for example, that dogs trained in agility tasks or Schutzhund protection work tended to be more persistent and curious, and to be more attentive to their owner (Marshall-Pescini, Passalacqua, Barnard, Valsecchi, & Prato-Previde, 2009; Marshall-Pescini, Valsecchi, Petak, Accorsi, & Previde, 2008). We might expect that dogs specifically trained to respond to verbal commands might be more likely to respond to their name, or more successful at doing so in noise. In order to explore whether our participants' performance differed based on either breed or working status, we combined the datasets from the two studies in which the dogs were successful at recognizing their name in noise (5 dB and 0 dB) to see if there are any Breed Group or Working Dog Status differences that lead to better performance in name recognition. The combined dataset has 40 dogs (22 male). We performed a 2x2x2x7 ANOVA (Item by Noise Level by Working Dog Status by Breed Group). Unsurprisingly, this combined dataset replicated the general pattern of results seen in the two data sets individually: we found an overall effect of Item ($F(2, 39) = 24.25$, $p < .001$), but no effect of Noise ($F(2, 39) = 1.558$, $p > .05$) and no interaction ($F(1, 39) = .693$, $p > .05$).

More importantly, we found that the interaction between Item and Working Status was marginal ($F(1,39) = 3.463$, $p = .07$). Overall, working dogs, which include police K9s, search-and-rescue dogs, therapy dogs, service dogs, and service-dogs-in-training, listened longer to their name (8.16 seconds) and less to the foil (5.89 seconds) than pet dogs, who listened to their name for an average of 7.69 seconds and the foil for an average of 6.63 seconds. This may be an indication that dogs that receive more intense training, in general, also are more selectively responsive to their name (perhaps as a result of that additional training).

There was no interaction between Breed Group and Item ($F(6, 39) = .98$, $p > .05$), although this may be the result of the small number of dogs in some breed groups. Regardless of whether dogs were bred primarily for companionship, hunting, guarding, or herding, they preferred their name to unfamiliar foils, and no group performed better at this than any other group, despite the fact that working dogs tended to be in the Herding or Sporting groups.

### *Overall discussion*

The current studies examined whether domestic dogs could recognize a particular, highly-familiar word (their name) when spoken by a novel talker, and under what conditions they could do so. Below we discuss each of the main findings and their implications.

First, we found that dogs were successful at recognizing their name even when spoken by an unknown voice. This suggests that dogs have the capacity to generalize their lexical knowledge across talkers. Although this ability has been shown previously in individual dogs, it has not been shown to be a general capability. For example, the Yorkshire terrier in Griebel and Oller's study of word learning in dogs was able to comprehend commands given by an unfamiliar voice (2012). However, this particular dog was also highly unusual, in that she knew names for over 200 different objects and could retrieve them on command. The present study demonstrates that the ability to recognize a familiar word and generalize that across voices is found more generally among typical pet and working dogs.

Additionally, dogs demonstrated that they could respond to an unfamiliar voice even if the apparent speaker is not present in the room. While previous fMRI studies have shown that dogs display a clear neural response to the human voice, the dogs in those studies could not demonstrate any behavioral response, as they were within the fMRI (Andics et al., 2016, 2014). Our study shows that generally, dogs will behaviorally respond to their name even if a sound source is not immediately clear. This has practical implications for working dogs, like search-and-rescue dogs, who may need to take commands from someone other than their handler in emergency situations, and may need to do so at a distance, when the speaker is out of view.

We also found that dogs could succeed at this task even in the context of multitalker background babble. When the noise was softer or at the same level as their name, dogs recognized and preferred to listen to their name over another dog's name. When the noise was louder than their name, dogs no longer showed that preference.

One question we asked was how similar dogs' performance at a hearing-in-noise task would be to infant performance. Infants tend to fail at this task when the signal and noise are at the same intensity, but dogs were very successful at this level. Since dogs were successful at this task, but do not have linguistic processing, they must utilize only domain-general auditory processing mechanisms for name recognition. It is possible that infants, too, rely on these domain-general mechanisms for this task as well.  Clearly, though, the fact that young infants have limited linguistic skills is, of itself, not sufficient explanation for their poor performance listening in noise.

Infants' performance could also be due to their poorer attentional capabilities or deficits in auditory processing. Identifying the extent to which attention, auditory processing, and linguistic knowledge contribute to comprehension of speech in noise is necessary to understand why infants have difficulty with this task.  Use of a domestic dog comparison group can highlight the separate contributions of the attention system and auditory system to speech-in-noise perception. By using dogs, we can control for linguistic prior knowledge. Future studies will compare dog and infant auditory and attentional capabilities to determine their similarities. This will allow further studies to compare dog and infant performance in listening-in-noise tasks to tease apart the attention and auditory system contribution.

We found hints that working dogs were performing better than pet dogs. While the effect was only marginal, it appears that working dogs showed both longer listening for their own name, and less listening to the foil name. One possibility is that these dogs hear their names more often than pet dogs. Perhaps when owners ask their pet dogs to perform tasks, they just state a command ("sit!") rather than specifying the dog first ("Tahoe, sit!"). Or perhaps pet owners are more variable in what they choose to call their dog, using a name in some instances and a nickname in others. In contrast, working dogs may hear just one name very often, a name that is very salient to the dog. Indeed, many search-and-rescue dog owners specified that their dog has a call name, which is consistently used while the dog is in the field, but while at home (and not working) the dog is more freely called both their name and also nicknames (as is the case for many pet dogs). It is also possible that the increased obedience and task training that working dogs receive leads to better overall attention abilities. This would lead to better attention in our task and better overall performance.

The results of the current studies have practical implications for the training and use of service dogs, search and rescue dogs, and other working dogs. Working dogs must contend with many different noisy environments. Cities, one common location for service dogs, tend to have ambient environmental noise at around 70 dB, which is five dB louder than average conversational speech (Appleyard & Lintell, 1972; McAlexander et al., 2015). Dogs in the current study fail to listen longer to their own name in noise as compared to other names at -5 dB SNR. This suggests that hearing target speech that is five decibels less intense than background noise may be at the limit of what dogs are capable of perceiving. This noise level should be kept in mind when dogs are working in the field.

In conclusion, the present study begins an exploration of dogs' speech perception abilities in noisy environments. The findings suggest that dogs are capable of understanding and attending to an unfamiliar voice both in quiet, and in the presence of competing distractor voices. Dogs are successful when the noise is softer or at the same intensity as the target speech; however, they fail to recognize their name when the noise is louder than the target. Future work will explore dogs' speech perception capabilities in more detail and provide comparisons with infant speech perception.

# Chapter 3: Attention and audition: Using a dog model to explore the underlying causes of information masking for infants hearing their name in single-talker background noise

## *Overview*

Infants often hear speech in the presence of one or more background talkers. The presence of background talkers significantly impacts infants' ability to understand words - even familiar words like the infants' own name (Newman, 2005). Speech comprehension with multiple talkers in the background is a more difficult task for infants than adults; in order to comprehend the target speech, infants require the speech to be significantly more intense than the background noise in comparison with adults (e.g., Newman, 2005, for infants; Rosen, Souza, Ekelund, & Majeed, 2013, for adults). While infants generally struggle with speech-in-noise perception in comparison to adults, infants are much more impacted than adults by the presence of a single talker in the background, as compared to multiple talkers in the background (Newman, 2009). Adults typically find it easier to perceive speech in the presence of a single talker in comparison to multiple talkers (Rosen et al., 2013). Infants, on the other hand, find it much harder to understand speech in a single-talker background (Newman, 2009).

There are two main reasons why infants may have more difficulty comprehending speech in the presence of a single background talker, as compared to multiple background talkers: difficulties detecting and segregating target speech from background speech, and immature sustained and selective attention (difficulty specifically attending to the target stimuli over time). No studies have examined the individual effects of infants' detection and segregation of speech streams as well as attention on speech perception in the presence of a single talker. It is difficult to differentiate these effects by only testing infants, as it is hard to isolate the individual effects of the auditory and attention system on speech-in-noise perception when these systems are so interconnected in their function.

One way to differentiate auditory contributions from attentional functions in a speech-perception-in-noise task is to compare infants' performance to that of domestic dogs. Studies have shown that domestic dogs have a mature peripheral auditory system similar to that of an adult human (Strain, 2012). However, their attention abilities in speech perception tasks are relatively similar to that of infants (see Mallikarjun, Shroads, & Newman, 2019). As a result, if dogs outperform infants in certain listening-in-noise tasks, it suggests that those tasks primarily rely on the auditory system. Conversely, if dogs perform similarly to infants, the tasks may be more attention-oriented.

In this chapter, I will first discuss the factors that affect infants' speech perception in noise. I will then discuss some research that suggests infants' difficulties in this task stem from both auditory and attentional difficulties. Lastly, I will describe why a domestic dog model can aid in experimentally demonstrating the extent to which auditory and attentional processing difficulties are responsible for infants' performance.

### *Auditory Processes*

In order to understand speech perception in noise, listeners must first perceptually segregate the target speech from the background noise. Infants' immature auditory system can lead to problems detecting and segregating target speech from the background, which can make stream segregation more difficult. While infants can use some cues to differentiate between streams, like the time at which the streams begin (Bendixen et al., 2015), they have difficulty utilizing other informative cues for stream segregation, like the location of sounds (Ashmead et al., 1987).

One potential reason why infants specifically struggle perceiving speech with a single talker in the background may be that stream segregation becomes much harder when the streams are very similar to one another. Even in adults, segregating similar sound streams can be difficult. A substantial release in masking can be observed when the target stream is manipulated to become less similar to the distractor stream through spatial separation, shifted onset times, and decreased target-masker similarity (e.g., Durlach et al., 2003; Kidd, Mason, Deliwala, Woods, & Colburn, 1994; Neff, 1995). Release from masking due to a reduction in target-masker similarity can also be seen with infants. Infants at 7.5 months can discriminate between an unfamiliar female voice and an unfamiliar male background talker at 10 dB SNR (Newman & Jusczyk, 1996); however, at the same SNR, 7.5 month-old infants struggle to distinguish between two unfamiliar female voices (Barker & Newman, 2004). When the target speaker and the background talker are similar to one another in spectral features, like fundamental frequency, and temporal features, like speech rate and onset time, it can be difficult to assign the perceived speech sounds to the correct source (see Darwin, 2008, for a review of speech-in-noise perception).

### *Attentional processes*

A portion of infants' attentional difficulty can be attributed to poor sustained attention, or the ability to maintain attention on a stimulus for a certain period of time. For example, 6-9 month-old infants' absolute thresholds for detection of tones in quiet and noise are poorer than adults; additionally, children's psychometric functions had shallower slopes than adults' psychometric functions, which indicates that as the signal intensity increased, infants showed a relatively slower increase than adults in accurate signal detection (Bargones, Werner, & Marean, 1995). This difference between infants and adults can be partially attributed to infants failing to behaviorally demonstrate that they detected the target sound on a higher proportion of trials. Infants may fail to demonstrate that they detected the target sounds due to difficulty sustaining attention on the task. It is also possible that infants are not attending to the auditory stimuli during a portion of trials in speech-perception-in-noise studies, and instead focusing on irrelevant visual stimuli. As such, what may be coded as infants attending to the stimulus may be infants turning to attend to other visual stimuli in the testing area, or turning due to fidgeting or fussiness.

Infants also can have difficulties comprehending speech in noise due to their developing selective attention abilities in the auditory domain. Auditory selective attention can refer to two different but related concepts: preference for listening to certain streams longer than other streams, even if they may not necessarily occur simultaneously, and attending to a specific target stream in the presence of a distractor stream (Gomes, 2000). While several studies have demonstrated that infants show preferences for certain sounds over others (mother's voice over stranger's voice, Decasper & Fifer, 1980; infant-directed speech over adult directed speech, Fernald, 1985; happy-affect speech over neutral-affect speech, Singh, Morgan, & Best, 2002), it is harder to specifically examine selective attention to one of two simultaneous streams, especially since infants cannot be specifically instructed to attend to one stream, and cannot report specific words or sounds. Evidence for infants' difficulty with selective attention to simultaneous streams primarily comes from psychoacoustic tasks with non-speech stimuli, where infants tend to listen broadly to multiple aspects of speech rather than narrowing in on the specific cues that provide the most information (Bargones & Werner, 1994). Infants fail to narrow their selective attention to focus on the most important aspects of sounds; as such, they may pay undue attention to irrelevant aspects of sound streams, and have difficulty focusing on the key elements of the target stream.

Studies of selective attention to frequency in infants suggest that infants have difficulty specifically attending to information within the frequency band of the expected auditory stimulus (Bargones & Werner, 1994; Werner & Bargones, 1991); this may be responsible for their difficulty perceiving speech in the presence of background talkers. For example, infants have difficulty ignoring background noise even when the background noise does not overlap in frequency with the target signal; infants habituated to the syllable /bu/ in quiet dishabituate when they hear the unfamiliarized syllable /gu/, but fail to do so when they were habituated to the syllable /bu/ in the presence of high frequency sounds, including cricket noises, which were at a higher frequency than the target syllables (Polka et al., 2008). When listening to speech in noise, infants' less-selective attention could lead to difficulty comprehending words.

### *Differentiating auditory and attentional effects in infant speech perception in noise*

Behaviorally teasing apart auditory and attentional contributions to difficulties with speech perception in noise is difficult, especially when only testing infants and adults. While adults have a mature auditory and attentional processes, both of these systems are immature in infants, so comparisons of adults and infants fail to identify which system is primarily responsible for infants' poorer performance. Additionally, the behavioral tasks used to test infants' speech perception in noise inherently involve both of these systems, as in these tasks, infants must segregate two streams of speech stimuli and selectively attend to the target stream containing a word of interest. As such, it is very difficult to identify the individual effects of the auditory and attention system on infants' difficulties understanding speech in the presence of a background talker.

Given the similarities and differences in auditory and attentional function in dogs and infants, comparing their performances in a similar speech-perception-in-noise task can separate contributions of these individual systems. In the auditory domain, dogs have better auditory perception abilities than infants, who are still developing higher-level auditory skills like stream segregation and sound localization (see Werner, 2007, for a review of infants' auditory skills). Adult dogs' auditory system is closer to that of adult humans' in their ability to distinguish between frequencies (Bach et al., 2016). Prior data has shown that dogs are better able to recognize their name in multitalker background noise than infants (Mallikarjun et al., 2019); this is a task that relies heavily on the auditory system to overcome energetic masking (Culling & Stone, 2017).

In terms of attention, while studies have not directly compared the dog attention system to the infant attention system, our lab has used similar stimuli in the same testing paradigm to examine dogs' and infants' levels of attention to auditory stimuli. It is important to note that while dogs and infants may have differing levels of attention for other stimuli (e.g., food), this dissertation is particularly concerned with attention to human speech and the effect of auditory distractors on speech perception. Our lab has found that while dogs overall look for less time at speech stimuli than infants, in name-in-noise studies, both dogs and infants show a similar proportional preference for their own name over the foil name; in studies where participants successfully show a preference for name over foil, dogs listen to their own name in 9-talker background noise about 1.23 times longer than the foil name in noise (Mallikarjun et al., 2019), and infants listen to their own name in 9-talker background noise about 1.21 times longer than the foil name in noise (Newman, 2009).

This chapter uses a dog model to explore whether primarily attentional or auditory factors lead to infants' particular difficulties perceiving speech with a single talker in the background. In order to assess whether attentional or auditory factors play a larger role in infants' difficulties understanding speech with a single-talker background, it is necessary to compare dogs' and infants' relative performances in nine- and one-talker background noise, rather than just comparing their performances in single-talker background noise alone. Dogs have a more sensitive auditory system in comparison to infants, and are capable of comprehending speech at a more difficult signal-to-noise ratio than infants, as seen in our previous study (Mallikarjun et al., 2019); as such, it is necessary to examine dogs' performance in single-talker background noise relative to their performance in nine-talker background noise. Comparing dogs' relative performance in single-talker and nine-talker background noise with existing data of infants' relative performance can suggest whether auditory or attentional factors are primarily responsible for difficulty with single-talker background noise.

### *Experiment 1: Dogs' Recognition of their Own Name in 0 dB SNR One-Talker Background Noise*

Mallikarjun, Shroads, and Newman (2019) showed that dogs can understand speech in the presence of nine background talkers. However, no studies have yet examined dogs' understanding of speech with a single background talker. To test whether dogs have an easier or more difficult time with single-talker background noise than infants, I used the same testing paradigm and similar stimuli as used in the prior single-talker infant study (Newman, 2009). Dogs' name recognition in single-talker background noise was compared to dogs' name recognition in 9-talker background noise from Mallikarjun, Shroads, and Newman (2019). This comparison will determine whether the pattern of dogs' name recognition is similar to adults, who can more easily recognize speech in single-talker noise than nine-talker noise, or infants, who can more easily recognize speech in nine-talker noise than single-talker noise.

If dogs find single-talker noise easier than nine-talker background noise, like adults, it suggests that the difficulty of single-talker background noise for infants may stem primarily from their underdeveloped auditory system. If instead, like infants, dogs find single-talker noise more difficult than nine-talker background noise, it suggests that single-talker noise is difficult primarily due to attentional factors that are shared across the two species.

**Participants.**

In this study we tested 20 adult dogs (12 M). For each participant, we collected information on age, length of time with name, breed, and working status (whether or not the dog has a job, such as bomb detection or service). Dogs were required to have their name for at least 10 months in order to participate in the study, and to be at least 1 year old. We excluded any dogs that were taking psychiatric medication, and dogs whose owners noticed any signs of hearing loss.

On average, the dogs were 4.25 years old, and had been hearing their name for 3.9 years. We categorized dogs' breeds by American Kennel Club breed group: Toy Group (3), Working Group (2), Sporting Group (5), Herding Group (3), Non-Sporting Group (1), and mixed breed (6). Four of these dogs were therapy dogs, and one was a service-dog-in-training.

Five additional dogs were tested, but were excluded due to equipment malfunction (2), prior familiarity with the foil name (1), failure to meet study requirements about length of time with name (1), and external noise disrupting the testing (1).

**Test materials.**

The stimuli for this study consisted of a target sound stream and a distractor sound stream. The distractor stream was a single female voice reading aloud from a book. This voice was one of the voices recorded previously by Newman (2005) to explore infant's perception of their name in nine-talker background noise. This speaker was used because her fundamental frequency and speech rate was perceptually similar to that of speaker for the target sound stream. This similarity suggests that this particular distractor stream voice would be one of the more difficult voices out of the nine voices recorded in Newman (2005) to differentiate from the target speaker. If dogs succeeded with this difficult voice, it is more likely that they would also be successful with the other background voices.

The distractor stream speaker recorded the passage using a Shure SM51 microphone in a sound-attenuated room. This stream was edited to remove any long pauses, like page turns or when the speaker needed to clear her throat. As a result, the passage sounded more like a monologue than conversational speech; however, it still sounded like natural speech. All of the dogs heard this talker for the distractor stream.

Prior to the study visit, dog owners were asked for the name or nickname that their dog is most commonly called. Each name was matched with a foil name. In order to prevent any bias caused by the speaker producing target names in a more excited manner, each foil was chosen from the existing set of recorded dog names, some of which were recorded for studies by Mallikarjun, Newman and Shroads (2019), and some of which were target names for other scheduled participants in this study. The foil was matched to the target name in the number of syllables and stress pattern, and the names were chosen to be as phonetically dissimilar as possible from the original name in phonemes (e.g. *Bruno* could be matched with the foil name *Sally*).

This study used only one foil, rather than the three foils used in the infant studies in Newman (2005). Additionally, this study presented the dog's name and the foil in both quiet and in noise, while the infant studies presented the infant's name in noise and three foils in noise, with no quiet trials. We have found that dogs do better in studies if they are presented with auditory stimuli that have variety; this may be because the presence of quiet trials makes the noise trials seem more interesting. Dogs also tend to disengage entirely from studies that they find too difficult (Mallikarjun, Shroads, and Newman, 2019). As a result, it was useful to add easily understood stimuli in the study, like the quiet name and quiet foil, so dogs did not disengage.

A female native English speaker from eastern Pennsylvania individually recorded the target stream, consisting of a name and foil name for each participant. The speaker was recorded saying the dog's name and the foil name in dog-directed speech with a variety of inflections and durations (for more information on dog-directed speech, see Ben-Aderet et al., 2017).

A total of fifteen tokens were selected out of the original recording of target names, matched to the fifteen tokens of the foil name file as closely as possible for pitch, duration, intonation contour, emotionality, and vocal quality. Pauses between tokens of dog names were adjusted such that the target and foil files had the same overall duration of 22 seconds. There was an initial silence period for 0.5 seconds.

The name and foil speech streams for each study were then mixed with the distractor speech stream to achieve a specific SNR. However, the names and foils in the target speech streams are separated by long pauses, while the distractor speech stream had previously been edited to remove any disfluencies or pauses, so there were very few periods of silence in the distractor stream. As such, before adjusting the average RMS amplitude of the target stream to achieve a specific SNR, the periods of silence were removed between the names and foils. The average RMS amplitude was adjusted using this target speech and the background stream to set the SNR. After doing so, the periods of silence were added back to the target streams.

In this study, the distractor was first adjusted to the same intensity as the target speech (0 dB SNR), which was the most difficult level at which dogs succeeded at preferring their own name to the foil name in nine-talker background noise (Mallikarjun, Shroads, and Newman, 2019).

For comparison, the lowest level for nine-talker background speech at which one-year-old infants are generally successful at name recognition is +5 dB SNR (Newman, 2005).

**Apparatus.**

The experiment took place in a six-foot by six-foot three-sided test booth made of pegboard. The walls of the booth were four feet high. A tan curtain hung from the ceiling to the top of the booth to ensure that the dog could not see over the booth. On the front wall of the booth, there was a hole cut out for the lens of a video camera. The video camera recorded each session and allowed the coder to see the dog’s behavior inside the booth. Above the camera, a light was mounted in the center of the panel. The side walls each had a light mounted in the center and a speaker directly behind the light to play stimuli for the dog. A Windows computer was used behind the front wall of the booth to run the experiment and to code the study. The experimenter used a keyboard to start trials and code the dog's looking behavior.

**Procedure.**

The dog and his or her guardian were brought into the booth by an experimenter and the guardian signed consent forms. The dog sat on the owner's lap or directly in front of the owner, depending on their size and what made them most comfortable. The dog's guardian was provided with headphones and masking music to prevent him or her from biasing the dog's responses. The dog either sat facing towards the front of the booth (towards the camera) or facing the back of the booth (towards the owner). In either case, the dog's attention was maintained as much as possible at a point equidistant from the two sides of the booth (where the loudspeakers were located). As a result, the dog's natural inclination upon hearing a sound through a loudspeaker was to turn their head or body 90 degrees to face the source of sound.

There were two practice trials of classical music to familiarize the dog with the procedure. Newman (2005, 2009) used more practice trials with infants, but dogs tend to lose interest in practice trials, so in this study, only two were used. Their listening time was judged by the amount of time they spent looking at the sound source (the wall behind which the speaker was mounted).

Two experimenters ran the test phase portion of the study; one to code the dog's looks (the coder), and one to produce auditory attention getters (the Attention experimenter). At the start of the test trials, the light on the front of the booth turned on, and the Attention experimenter made a sound to get the dog's attention to the front of the booth. The sound was used to attract the dog's attention; although work with infants typically uses only lights, pilot work suggested that the light alone was not a sufficient attention-getter for most dogs. Lights are still used for the dog studies because they help the coder code the dogs' looks to the correct sound source. Once the dog attended to the front, a light turned on in either the left or right side of the booth. The Attention experimenter then made a sound on that side. Once the dog attended to that side, the stimulus played from the speaker on that side. The coder used a keyboard to code the dog's looks towards and away from the sides. The stimulus played for a full twenty-two seconds or until the dog looked away for two consecutive seconds, whichever occurred first. Any time the dog spent looking away was subtracted from the dog's overall looking time. The coder wore Peltor aviation headphones playing masking music so she would not be able to hear the trials and have that influence her coding.

**Results.**

This study tested dogs on name preference in quiet and in single-talker background noise at 0 dB SNR, where the target speech is the same intensity as the background speech. In a prior study, dogs preferred their name to another dog's name in 0 dB SNR when the background speech contained nine talkers (Mallikarjun et al., 2019); comparing this one-talker background study with the previous nine-talker background study at the same SNR will allow us to determine which type of background noise is more difficult for dogs. Dogs' relative results in single- and nine-talker background noise could then be compared to patterns of results seen in adults and infants.

***Dogs' name preference in quiet and single-talker background noise at 0 dB SNR***

On this 0 dB SNR single-talker dog data, I ran a 2x2 ANOVA with Noise (Quiet versus Noise) and Item (Name versus Foil) as within-subject factors. There was a significant main effect of Noise, $F(1, 19) = 6.535$, $p < .05$, where dogs listened, on average, longer to trials in quiet (8.46 seconds) than trials in noise (6.86 seconds). There was no main effect of Item, $F(1, 19) = 2.32$, $p > .05$, meaning there was no significant difference between mean listening times to the dogs' own name and the foil name. There was also no interaction between Item and Noise, $F(1, 19) = 1.904$, $p > .05$. Despite the lack of interaction, I compared dog's average listening time to their name and the foil in both quiet and in single-talker background noise. A paired samples t-test showed that in quiet, dogs preferred to listen to their name in comparison to the foil, $t(19) = 2.12$, $p<.05$. This was somewhat expected, as multiple prior studies have shown that dogs prefer their name to a foil name in quiet (e.g., Mallikarjun et al., 2019). A paired samples t-test showed that in noise, there was no difference between the means of dogs' listening times to their name in single-talker background (mean = 6.79 seconds) and the foil in the same background noise (mean=6.92 seconds), $t(19) = -0.18$, $p > .05$. Dogs' preference for their name to the foil in quiet suggested that the negative result in noise is not the result of dogs disengaging with the study itself, but the result of dogs' difficulty recognizing their name in the single-talker background noise at 0 dB SNR.

***Dogs' name preference in noise only (single- and nine-talker background noise at 0 dB SNR)***

I next ran a 2x2 ANOVA on only the trials from both this study and Mallikarjun et al. (2019) that occurred in noise (either 9-talker or single-talker) with Item (Name versus Foil) as a within-subject factor and Number of Voices (Nine versus One) as a between-subjects factor. Figure 12, below, shows two graphs displaying the data from this study and from Mallikarjun et al. (2019). In this noise-only dataset, there is no main effect of Voices ($F(1, 38) = .39$, $p > .05$) or Item ($F(1, 38) = 1.99$, $p > .05$). There is a marginal effect for the interaction between Voices and Item, $F(1, 38) = 3.032$, $p = .08$.

Mallikarjun et al. (2019) showed that there is a significant difference between Name in Nine-Talker Background Noise and Foil in Nine-Talker Background Noise, $t(19) = 2.27$, $p < .05$; this result indicated that dogs preferred their name (mean = 6.85 seconds) to the foil (mean = 4.93 seconds) in nine-talker background noise at 0 dB SNR. As discussed above, I had also performed a t-test examining Name in One-Talker Background Noise and Foil in One-Talker Background Noise, $t(19) = -0.18$, $p > .05$, and found no difference between the means of dogs' listening times to their name in single-talker background (mean = 6.79 seconds) and the foil in the same background noise (mean= 6.92 seconds).

Together, these results suggest that dogs have more difficulty recognizing their name in single-talker background noise than in 9-talker background noise. This is an interesting finding. Dogs in this experiment showed no effect of Item in 0 dB SNR single-talker background noise, while the dogs in the 9-talker background version of this study *did* show an effect of Item at 0 dB SNR. Dogs have more difficulty recognizing their name in single-talker background noise than in 9-talker background noise; this is the same pattern seen in 5- and 8.5-month-old infants tested by Newman (2009), and opposite of the pattern seen in adult humans (Rosen et al., 2013).

Dogs have a much more sensitive auditory system in comparison to infants; this should mean that dogs' stream segregation abilities are also more precise than that of infants. However, dogs still find single-talker background speech more difficult than nine-talker background speech. Why would dogs show the same pattern as infants, given their more sensitive auditory system? Perceiving speech in the presence of a single talker requires both precise stream segregation and attentional capabilities. Dogs' selective and sustained attention has been shown to be similar to that of infants in name-perception-in-noise tasks (Mallikarjun, Shroads, and Newman, 2019); this suggests that the difficulty with single-talker background speech for both infants and dogs is not a stream segregation problem, but instead an attentional problem.

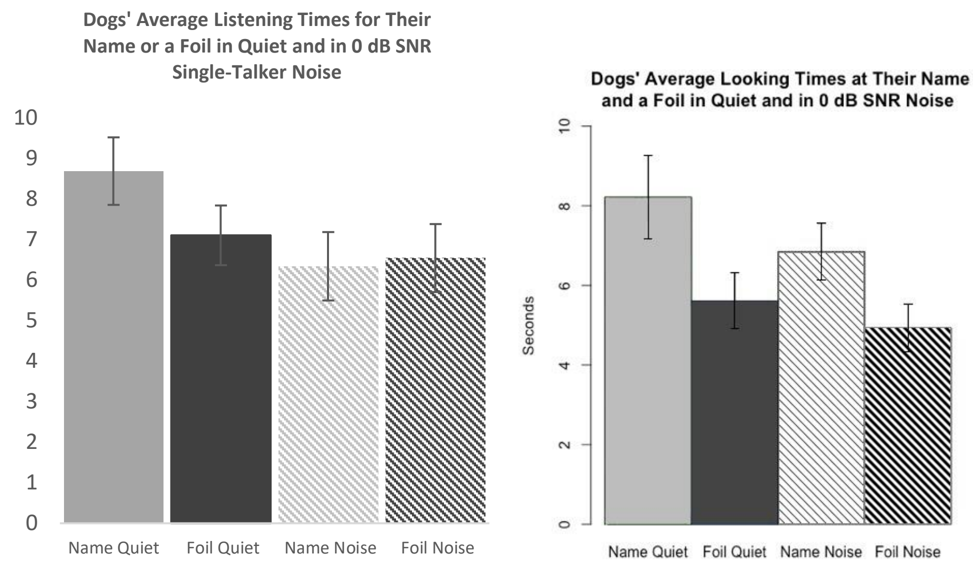


Figure 12: The figure on the left is a graph of dogs' listening times in 0 dB Single-Talker Noise. The figure on the right is a graph of dogs' listening times in 0 dB Nine-Talker Noise. While dogs do not successfully prefer their name to the foil name in the single-talker experiment, they do so in the nine-talker version (Mallikarjun et al., 2019).

***Working dog analysis***

This study provided some evidence that attentional difficulties may be the cause of infants' difficulties perceiving speech in noise through use of a comparative dog model. This additional analysis specifically compares pet dog performance to working dog performance to better assess whether attentional processes play a significant role in dogs' difficulties recognizing their name in single-talker background noise. Understanding the underlying cause of dogs' difficulties allows for a better comparison to infant performance in the same task.

A successful comparative model demonstrates *specificity* (the model shows the behavior in question) as well as *generality* (the results from the model would be informative about human behavior). A dog model shows specificity for speech perception in noise, as dogs attend to speech in noise and like infants, prefer to listen to their name in comparison to a foil. The question of generality is more difficult, especially in terms of attention; for this study, I suggest dogs' auditory system is more similar to that of adult humans, and dog attention is closer to that of infant humans. However, both dog and infant auditory attention is underexplored. While we have shown similar performance in tasks that require auditory selective and sustained attention (Newman, 2009; Mallikarjun et al., 2019), other aspects of attention and general cognition may also play a role in speech perception performance in noise. One way to isolate the role of attention is by comparing populations that have a known difference in attention, but are similar in other cognitive domains.

Age has been shown to have an effect on dogs' reported attentiveness (Lit, Schweitzer, Iosif, & Oberbauer, 2010; Vas, Topal, Pech, & Miklösi, 2007); adult dogs and older dogs were rated as more attentive than juvenile dogs (Vas et al., 2007). Studies from our lab corroborate the findings from prior studies on age, as juvenile dogs tend to look less long to speech stimuli than adult dogs, even when it contains their own name. This lowered attention for speech is one reason why juvenile dogs are excluded from our studies. Since we already exclude young dogs from our studies, age is not included in this analysis.

In prior studies, dogs' training level has also been shown to have an effect on attentiveness; dogs that had received more training were rated as more attentive than dogs with less training (Lit et al., 2010; Vas et al., 2007). Additionally, Mallikarjun et al. (2019) demonstrated that dogs with more training (e.g., search-and-rescue dogs, therapy dogs, and service dogs) showed marginally stronger preference for their own name over another dogs' name. As such, comparing pet dogs with working dogs, who have higher levels of training than pet dogs, thus potentially better attention, could provide more evidence for the idea that speech perception in single-talker background noise is difficult primarily due to attentional demands. If in our studies working dogs show longer looking times than pet dogs, especially for their own name in noise, it suggests that attention plays a significant role in dogs' task performance. If working dogs and pet dogs perform similarly, it instead suggests that the reason dogs perform poorly in this task may have less to do with attention and more to do with other cognitive and behavioral issues, such as impulsivity, hyperactivity, and impulse control (Vas et al., 2007).

In this analysis, I compared working and pet dogs' performance in this study with single-talker 0 dB background noise with previously collected data of dogs' performance in single-talker 5 dB background noise, as well as dogs' performance in 9-talker 0 and 5 dB background noise from Mallikarjun et al. (2019). Information about these studies is in the table below (Table 1). I added the additional SNR levels to ensure that there are enough working dogs to accurately compare their performance to that of pet dogs. Given working dogs' increased training, I expected to see a difference in name recognition in noise between pet dogs and working dogs.

Table 1

| Study | Overall # of Dogs | # of Working Dogs | Average Age |
|---|---|---|---|
| Single Talker 0 dB | 20 | 4 | 4.25 |
| Single Talker 5 dB | 20 | 5 | 4.25 |
| Nine Talker 0 dB | 20 | 14 | 5.3 |
| Nine Talker 5 dB | 20 | 4 | 4.37 |

I conducted a 2x2x2x2x2 mixed ANOVA with Noise (Quiet versus Noise) and Item (Name Versus Foil) as within-subjects factors, and Working Status (Working versus Non-Working), Talkers (One versus Nine Talkers), and SNR (Zero versus Five dB SNR) as between-subjects factors. I included Noise x Item x Working Status as an interaction to determine whether working dogs show a greater preference for name over foil when they are listening in noise in comparison to pet dogs. SNR and Voices were not used in this interaction because there were not enough working dogs in each condition for this interaction to be of interest.

Prior to running the ANOVA, I used Levene's test of equality of variances to assess whether the variables and interactions of interest (specifically, Noise, Item and Working Status) had equal variance, given that there are fewer working dogs than pet dogs in the data set. Levene's test showed the variances of these variables were not different, $F(7, 312)= 1.04$, $p > .05$. Inclusion of the additional variables, SNR and Voices, also leads to Levene's test showing no difference in variance of the variables, $F(31, 288) = .81$, $p > .05$.

There were no main effects of SNR ($F(1, 76) = .869$, $p > .05$) or Talkers ($F(1, 76) = .615$, $p > .05$. There was also no main effect of Working Status, $F(1, 76) = .532$, $p > .05$, such that working dogs did not look longer overall at items in comparison to pet dogs. There was a main effect of Noise, $F(1, 78) = 7.731$, $p < .01$, such that dogs listen longer to trials in quiet as opposed to trials in noise. There was also a main effect of Item, $F(1,78) = 30.396$, $p < .001$, such that dogs listen longer to their name in comparison to the foil.

There was an interaction between Item and Noise, $F(1, 72) = 4.26$, $p < .05$. This was expected, since dogs overall preferred their name to the foil in quiet, but had more difficulty doing so in noise.

There was no interaction of Item, Noise, and Working Status, $F(1, 78) = 0.016$, $p > .05$, meaning that Working Status does not mediate the interaction between Item and Noise. I had expected that working dogs would be relatively similar to pet dogs in quiet, but show a stronger preference for name over foil in noise; however, this did not occur. There was additionally no interaction of Noise and Working Status, $F(1,78) = 0$, $p > .05$, such that working status did not affect overall listening times to items in noise or in quiet.

There *was* an interaction of Item and Working Status, $F(1, 78) = 4.796$, $p < .05$. Working dogs showed a stronger preference for their own name (mean looking time = 8.5 seconds) over the foil name (mean looking time = 6.46 seconds) in comparison to pet dogs' looking time to their name (mean = 7.5 seconds) and the foil (mean = 6.6 seconds). (see Figure 13, below). This is an interesting finding; while both pet dogs and working dogs recognize their name and prefer it to the foil, working dogs are better able to maintain their attention on their name. Working dogs' better performance in this task suggests that this task has a significant attentional component that isn't mediated by hyperactivity or impulsivity.

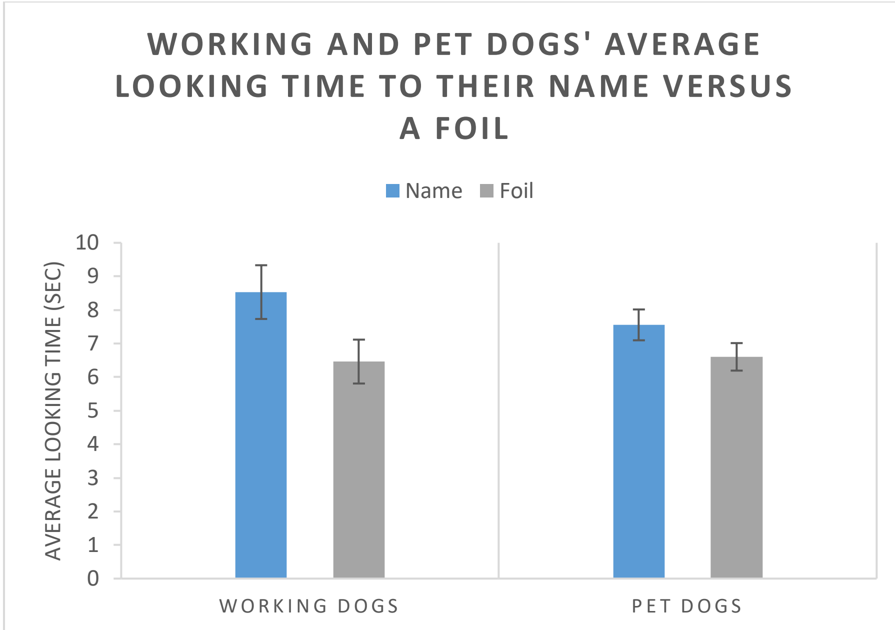


Figure 13: Working dogs show a greater preference for their own name over a foil than pet dogs.

### *Overall discussion*

This study used a comparative dog model to assess whether immature auditory processing or immature auditory attention is primarily responsible for infants' added difficulties perceiving speech in the presence of a single background talker, as compared to nine background talkers. Dogs and infants have similar attentional capabilities, but dogs have a mature, adult-human-like auditory system in contrast to infants' developing auditory system. As such, if dogs, like infants, struggle to understand speech in the presence of a single background talker, it would suggest that the task is especially difficult primarily due to attentional factors; since dogs have a sensitive auditory system more similar to that of an adult human's, if the task was especially difficult due to auditory factors, dogs would not struggle with the task.

Dogs do perform more like infants, showing a greater difficulty perceiving speech in a single-talker background than a nine-talker background at the same signal-to-noise ratio. This suggests that immature attention is the primary cause for the added difficulty perceiving speech in the presence of a single talker; despite dogs' mature auditory system, they still failed to show a preference for their name over a foil with a single talker in the background.

In order to accurately draw conclusions about speech perception in noise in infants from comparative models like dogs, it is necessary to understand auditory processes and attentional processes in both infants and dogs and how they compare to one another. One limitation of this study's comparison between dog and infant speech perception is that both dog and infant attention, especially in the auditory domain, have not been well explored. As such, it is difficult to separate the effects of other related cognitive processes, like impulsivity, inhibition control, and working memory from the effects of sustained and selective attention. Further studies in both dog and infant attention should attempt to tease apart these different factors to help isolate the effects of specific attentional factors on speech-in-noise perception.

Based on prior studies that showed that dogs' training levels were correlated with attention, but not with impulsivity or hyperactivity (Vas, Topal, Péch, & Miklósi, 2007), a further analysis was done on a larger dataset comparing the speech-perception-in-noise performance of highly-trained working dogs with pet dogs. If the dogs with more training found it easier to understand speech in the single-talker background noise than dogs with less training, it would suggest that attentional factors play a large role in the increased difficulty of speech perception in single-talker background noise. Working dogs looked longer overall at their own name than a foil name in comparison with pet dogs. This result provides further evidence that the added difficulty perceiving speech in single-talker background noise is due to an increased attentional burden, and variability in dogs' hyperactivity or impulsivity is not necessarily responsible for dogs' poor name recognition in single-talker background noise in our task.

However, the prior evidence for working dogs having better attention has some limitations, in that these previous studies mostly used owner questionnaires to assess attention, hyperactivity, and impulsivity. While these questionnaires were adapted from well-tested human questionnaires that assess ADHD (DuPaul, Power, Anastopoulos, & Reid, 1998), further behavioral studies in dogs that separate attentional factors from hyperactivity and other cognitive factors are necessary to accurately assess whether dogs' additional difficulty understanding speech in a single-talker background as compared to a nine-talker background is specifically due to poor attention, or whether a portion of dogs' poor performance can be attributed to other cognitive processes.

Further studies can test infants with different types of background noise to examine which acoustic and linguistic features of the background noise lead to the greatest amount of distraction (see Chapter 4). Additionally, experiments can be conducted using more ecologically valid noise to determine how commonly occurring auditory distractors can impact infants' word recognition.

While this comparative research was primarily meant to help identify the dominant systems underlying speech perception in noise in infants, the results from the dog model can also provide some practical guidance for dog owners and handlers. Service dogs and other working dogs are often in situations with varying levels and types of background noise, from malls and restaurants to parks and busy streets. Identifying the types of noise that most impact dogs' speech perception can help handlers understand why their dog might be less responsive in certain situations, and allow them to find better ways to work with their dogs in difficult environments.

Overall, this study provides some evidence that infants' added difficulty perceiving speech in a single-talker background as compared to a nine-talker background is primarily due to an increased attentional burden. Further studies are necessary to determine the aspects of single-talker speech that make it particularly attentionally challenging.

# Chapter 4: Infants' perception of their name in single-talker background noise: Effects of temporal modulation and presence of comprehensible speech

## *Overview*

Infants have much more difficulty understanding speech in the presence of a single background talker as compared to adults (Newman, 2009). The previous chapter uses a domestic dog model to determine whether auditory system deficits or attentional deficits play a larger role in infants' poorer performance understanding speech with a single talker in the background. The results indicate that dogs find single-talker noise to be more difficult that nine-talker background noise, which is the same pattern seen in infants. Dogs show difficulty understanding speech in single-talker background noise despite their more sensitive auditory system; this provides evidence for the idea that that single-talker background noise is difficult for infants (and dogs) primarily due to attentional demands.

There are several potential sources for the increased attentional demands on listening to speech in single-talker background noise. Understanding speech in single-talker background noise requires the listener to selectively attend to the target speech while ignoring or attenuating the background speech. This task is especially difficult because the background speech, a single talker, is acoustically similar to the target speech. This can cause several problems for the listener. First, the similarity between the target and distractor cause source confusion for the listener, where the listener has trouble identifying which stream is the target stream. Second, there are several aspects of a single-talker background stream that can cause distraction for the listener, which may lead to additional difficulty focusing on the target stream. These aspects include the comprehensibility of the distractor and the temporal variation within a single talker distractor stream, both of which can draw attention away from the target stream.

This chapter will specifically focus on the aspects of a single-talker background stream that cause distractibility. The goal of this chapter is to specifically examine which of two potential components that lead to distraction, temporal variation of speech, or the presence of comprehensible speech, contributes more to infants' difficulties understanding speech when listening in noise.

Certain features of background noise can draw attention away from the target speech and distract the listener. First, infants may be distracted by the temporal, or time-based, properties of the background speech. There are few infant studies that explore the effect of amplitude modulation on distraction during speech perception. One study with seven-month-olds found that infants' ability to discriminate between vowels was worse in the presence of an amplitude-modulated masker than a constant masker (Werner, 2013). In single-talker speech, there is much more amplitude fluctuation than in multitalker speech, due to the natural breaks and pauses between words; as such, increased amplitude modulation may be one reason why infants find speech comprehension in single-talker background noise difficult.

Infants also may be distracted by the presence of a comprehensible voice in the background. Infants show an early bias for speech sounds in comparison to other sounds (Vouloumanos et al., 2010; Vouloumanos & Werker, 2004); this early bias could make it easier for infants to learn language (Vouloumanos & Werker, 2007). As a result of their general preference for speech, infants may be more distracted by background speech when they are listening to a target speaker because the background speech is also identifiably human speech. While in 9-talker background speech it is not possible to distinguish individual voices, a single-talker background is easily identifiable and could be a potential source of new information for the infant.

In addition to the potential distraction caused by a comprehensible voice, infants could experience phonological interference when familiar sounds from the background speech impede infants' ability to either identify the target speech or pay attention to the target speech. Infants develop native phonological representations within the first year of life (Cheour et al., 1998; Kuhl et al., 1992); as a result, infants could potentially detect their native-language phonology in the background speech, and that might be distracting for them. There are no studies in infants that directly explore this phenomenon. In preschool children, one study explored the effect of background sentences on an image identification task (Newman, Morini, Ahsan, & Kidd, 2015). Children were presented with target sentences that asked them to a point to a picture of a specific item in an array. These sentences appeared either in quiet, or in the presence of one of three kinds of distractor sentences. The distractor sentences consisted of standard sentences (e.g., *Airplanes fly quite high and fast*), jumbled sentences that contained all the words of their standard counterparts but in a nongrammatical order (e.g., *High. Fast. Fly. Airplanes. And. Quite.)* and reversed sentences that reversed the audio signal of the original sentence. Preschool-aged children performed more poorly in the presence of maskers than in quiet. Of the three types of maskers, children performed better in the presence of reversed sentences, and did equally badly when the maskers were normal sentences and jumbled sentences.

The fact that jumbled sentences were equally as disruptive as the normal sentences suggests that if preschool children are experiencing interference, it is not at the sentence comprehension level; rather, it is a result of lexical interference, phonological interference, or difficulty identifying the correct target auditory source. In infants, sentence-level and lexical interference are not likely causes for their difficulties perceiving words in background speech, since they do not know many words; they would only experience lexical interference if the background speech contained some of these early-learned words. While these factors have not been differentiated in this study or any other study, it leaves open the possibility that phonological interference may cause some difficulties understanding speech in the presence of background talkers.

Few prior studies have explored the individual effects of amplitude modulation and comprehensible speech. Newman (2009) tested infants on their perception of their own name with a time-reversed single talker in the background. Reversing a single talker creates a stream with the same spectral properties and similar temporal properties, but it no longer sounds like comprehensible speech. The study found that 5-month-old infants do not prefer their name to foil names in the presence of a time-reversed single-talker background. Young infants' performance in the reversed speech resembles their performance in the forwards, comprehensible single-talker background (see Figure 14). This finding could suggest that infants are not as distracted by comprehensible words in the background as they are by spectral and temporal properties within the background speech. However, it is also possible that infants were, in fact, distracted by comprehensible words in the forward speech, and the reversed speech was distracting for a different reason. Reversed speech sounds unnatural and violates many phonotactic rules; this alone could have drawn infants' attention. Despite the findings of Newman's initial study, the features of single-talker background speech that make it particularly difficult for infants are still unclear.

This proposed study will determine the extent to which temporal features of speech lead to difficulties in speech-in-noise perception for 13-month-old infants. I will examine infants' preference for their name in three different types of background noise that have been derived from the single-talker background speech used in the previous study. Infants will hear their name and a foil name in the presence of a single background talker, steady-state speech-shaped noise, and temporally-modulated speech-shaped noise that contains the same temporal variation as a single talker, without the comprehensible speech. Their performance in each of these types of background noise will be compared to determine to what extent amplitude modulation contributes to infants' poor performance in single-talker background noise.

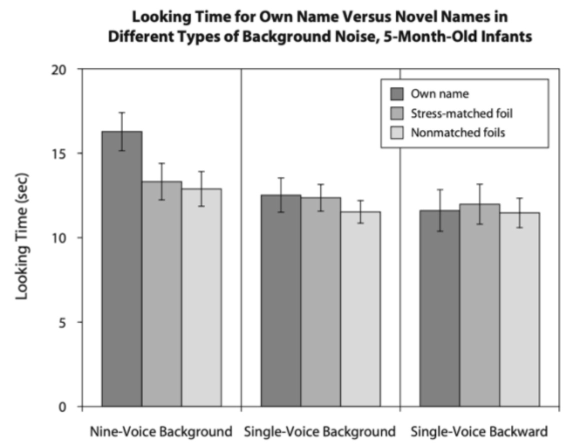


Figure 14: A graph from Newman (2005) showing infants' average looking times to their own name and foils in a nine-voice background, single-voice background, and single-voice backward (reversed) background. While 5-month-old infants prefer their name to the foils in the nine-voice background, they do not show this preference in the single-voice background or the single-voice backward background.

**Participants.**

The target number of participants for this study was 20 infants. However, due to the COVID-19 pandemic, it was only possible to collect data from 10 participants. I have also included data from four pilot participants that only heard 2 of the 3 possible trial types.

For each participant, data was collected on maternal education, language exposure, and ethnicity. This is collected primarily to ensure that this participant group is similar in demographic composition to demographic composition of Newman (2009).

**Materials.**

Stimuli consisted of a target sound stream and a distractor sound stream. There were three types of distractor streams: a Single Talker (ST), a Constant Amplitude Speech-Shaped Noise (SSN), which was constant-amplitude noise edited to contain the same spectral features as the ST distractor stream; and an Amplitude-Modulated Speech-Shaped Noise (Amp-Mod SSN), which consisted of speech-shaped noise with the same intensity modulations as the single talker.

The Single Talker distractor stream was a single female voice reading aloud from a book. This voice was one of the voices recorded previously by Newman (2005) to explore infant's perception of their name in nine-talker background noise. The same distractor stream was used in Chapter 3.

The amplitude modulation and spectral features of the other distractor noise streams were based on the amplitude modulation and spectral features of this single talker speech stream. To create speech-shaped noise based on this speech stream, a MatLab script was used (Nike, 2016). First, the spectral components were derived from the stream using Fourier transformation. The spectral phase of these components was then randomized. A reverse Fourier transformation was performed, yielding noise whose spectrum was nearly identical to the original speech stream. The resulting noise was the Constant Amplitude Speech-Shaped Noise (SSN) used in the study. This SSN shares the same amplitude and spectral features as the Single Talker stream.

To create the Amplitude-Modulated stream, I first used Praat (Boersma, 2001) to determine the intensity contour of the original Single-Talker speech stream. I then modulated the SSN in Praat using the intensity contour to create Amp-Mod SSN, which shares the same temporal features as the original speech, but lacks the linguistic content.

The target speech stream was recorded and the files were mixed and adjusted for SNR in the same manner as Chapter 3.

**Apparatus.**

The experiment took place in a six-foot by six-foot three-sided test booth made of pegboard. The walls of the booth were four feet high. A tan curtain hung from the ceiling to the top of the booth to ensure that the infant could not see over the booth. On the front wall of the booth, there was a hole cut out for the lens of a video camera. The video camera recorded each session and allowed the coder to see the infant's behavior inside the booth. Above the camera, a light was mounted in the center of the panel. The side walls each had a light mounted in the center and a speaker directly behind the light to play stimuli for the infant. A Windows computer was used behind the front wall of the booth to run the experiment and to code the study. The experimenter used a keyboard to start trials and code the infants' looking behavior.

**Procedure.**

The infant and his or her guardian were brought into the booth by an experimenter and they signed consent forms. The guardian was seated in the center of the booth and the infant was seated on their lap such that the participant was visible on the camera and was also equidistant from the speakers and lights. The equidistance from the speakers on the two sides of the booth was intended to prevent any bias for one side over another. By having the infant located in line with the two lights, each light was 90 degrees to the side of an infant facing forwards; this encouraged infants to turn their head fully to look at the source of sound, rather than just moving their eyes, which made coding the child's behavior easier. The guardian was provided with headphones and masking music to prevent him or her from biasing the child's responses. The experimenter also wore Peltor aviation headphones playing masking music so she was not able to hear the trials and have the trial sounds influence her coding.

The experiment was run using the BITTSy experimental testing platform (Newman et al., 2019). During the experiment, the child saw a light in the front of the booth begin to flash. Once he or she was looking center towards this light, a light on one of the sides of the booth flashed (the side was chosen randomly by the experimental program). Once the child looked at this side light, the stimulus played. The coder used a keyboard to indicate when the child turns his or her head at least 45 degrees towards the side. The trial typically ended when the child looked away for two seconds consecutively (Hayashi, Tamekawa, & Kiritani, 2001; Krumhansl & Jusczyk, 1990) – the assumption was that 2 seconds indicated actual inattention, whereas a brief glance away may not. If the child looked for the entire duration of the stimulus, the trial ended once the stimulus finished playing. The child was presented with three blocks of six trials (Name in Single-Talker Background, Foil in Single-Talker Background, Name in Amplitude-Modulated Speech-Shaped Noise, Foil in Amplitude-Modulated Speech-Shaped Noise, Name in Constant-Amplitude Speech-Shaped Noise, and Foil in Constant-Amplitude Speech-Shaped Noise) for a total of 18 trials. The order was randomized by block, such that the six types of trials appeared in each block, but each block had a randomized presentation of the six types of trials.

**Analysis.**

First, a preliminary analysis was done of the 10 participants that heard all 3 types of noise using a within-subjects 2x3 ANOVA with Name (Name or Foil) and Background Noise Type (Single Talker Background, Amplitude-Modulated Speech-Shaped Noise, or Constant Amplitude Speech-Shaped Noise) as factors. There was no main effect of Name, $F(1, 9) = 0.62$, $p > .05$. There was no main effect of Noise, $F(2, 18) = 0.56$, $p > .05$. There was no interaction between Name and Noise, $F(2, 18) = .11$, $p > .05$. Figure 15, below, shows a graph of the difference between Name and Foil for the three background noise types.

This null result with only 10 participants was not unexpected; similar studies with infants have tested over double this amount of participants to see results, with 20-36 infants per condition (Newman, 2005, 2009; Newman & Jusczyk, 1996).

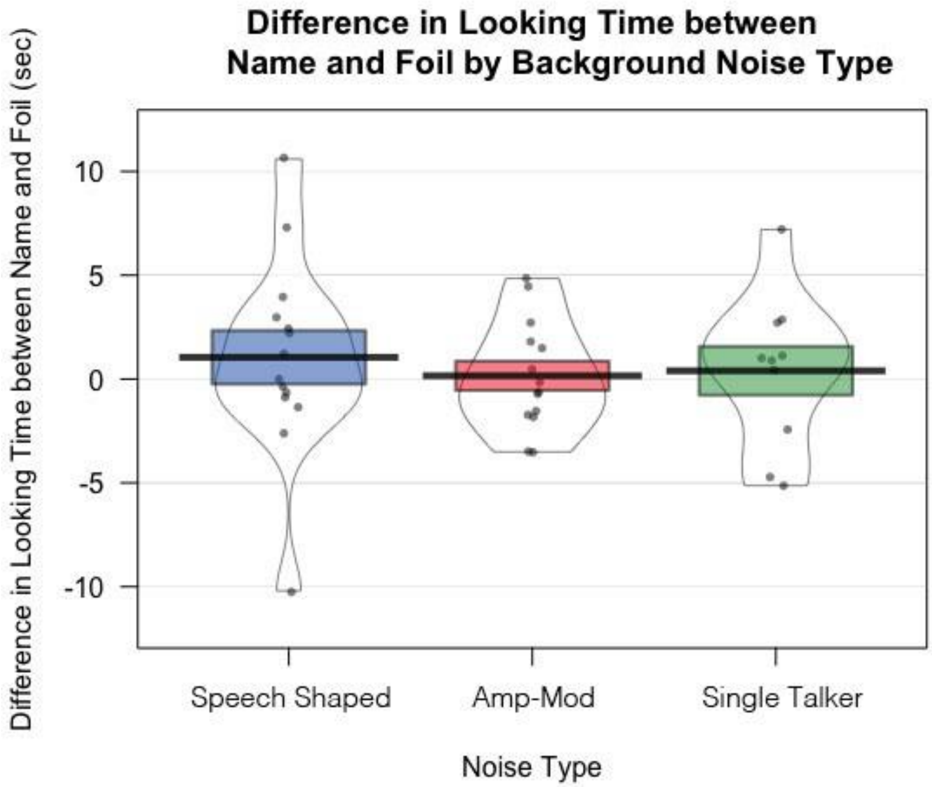


Figure 15: A graph showing the difference in looking time between name and foil in Speech-Shaped noise, Amp-Mod noise, and Single Talker Background noise. The colored bars show standard error around the mean, and the beans show the distribution of difference in looking time in each condition. There is no significant difference between the noise conditions in terms of difference in looking time to the name and foil.

An additional analysis was done with the 10 participants that heard 3 types of noise as well as the 4 pilot participants that only heard two types of noise: speech-shaped and amplitude-modulated. Instead of averaging looking time over the trial types, which is usually done in similar infant study analyses (Newman, 2005, 2009), I included data from all of the trials that each infant completed. Because some data is missing, an ANOVA was an unsuitable analysis. I used R (R Core Team, 2018) and the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) to fit a linear mixed-effects model using maximum likelihood to examine the individual effects of Name and Background Noise Type, as well as the interaction between Name and Background Noise Type. As random effects, I used intercepts for Participant and Block (the block of the study in which the trial appeared). Residual plots were used to check for deviations from homoscedasticity and normality, and visual inspection revealed no deviations. To determine whether Name, Background Noise Type, or the interaction between Name and Background Noise Type as fixed effects are significant, I used likelihood ratio tests to compare iterative versions of the model with and without each fixed effect and the interaction (see table).

First, a baseline model with no fixed effects and Participant and Block as random intercepts was compared to two models: one with Name as a fixed effect, and one with Noise as a fixed effect. The likelihood ratio test indicated that there is no significant effect of Name, ($X^2(1) = 0$, $p > .05$) and there is no significant effect of Noise ( $X^2(1) = 1.443$, $p > .05$). Table 2, below, shows these models and their corresponding $X^2$ values.

Table 2

| Model | Formula | $X^2$ | Significance |
|---|---|---|---|
| Baseline | LookingTime ~ 1 + 1\|Participant + 1\|Block | | |
| Name as Fixed Effect | LookingTime ~ Name + 1\|Participant + 1\|Block | 0 | p>.05 |
| Noise as Fixed Effect | LookingTime ~ Noise + 1\|Participant + 1\|Block | 1.443 | p>.05 |

Next, I compared a model with both Name and Noise as fixed effects to a model with only Noise as a fixed effect and a model with only Name as a fixed effect.

The likelihood ratio test indicated that the addition of both of these factors does not significantly improve the model in comparison to models with only one of these factors (Only Noise, $X^2(1) = .021$, $p>.05$; only Name, $X^2(1) = 2.064$, $p>.05$) (see Table 3).

Table 3

| Model | Formula | $X^2$ | Significance |
|---|---|---|---|
| Name and Noise as Fixed Effects | LookingTime ~ Noise + Name + 1\|Participant + 1\|Block | | |
| Name as Fixed Effect | LookingTime ~ Name + 1\|Participant + 1\|Block | 2.064 | p>.05 |
| Noise as Fixed Effect | LookingTime ~ Noise + 1\|Participant + 1\|Block | 0.021 | p>.05 |

Lastly, I compared the model with both Name and Noise as fixed effects to a model with the interaction between Name and Noise as a fixed effect to assess the significance of the interaction. This interaction is not significant, $X^2(2) = 1.588$, $p > .05$; Table 4, below, lists the models and their significance.

Table 4

| Model | Formula | $X^2$ | Significance |
|---|---|---|---|
| Name and Noise as Fixed Effects | LookingTime ~ Noise + Name + 1\|Participant + 1\|Block | | |
| The interaction between Name and Noise as a Fixed Effect | LookingTime ~ Name*Noise + 1\|Participant + 1\|Block | 1.588 | p>.05 |

Together, the linear mixed-effects model shows that neither Noise Type or Name have a significant effect on looking time for these 14 infants, and there is also no significant effect of the interaction between Noise Type and Name.

### *Discussion*

This study was intended to assess whether amplitude modulation or comprehensible speech content is the primary cause of informational masking when infants listen to speech in a single-talker background. Only 10 participants completed the full study, and an additional 4 participants participated in a version of the study that only examined the effect of amplitude modulation and not presence of comprehensible speech. The preliminary results with these 14 participants do not provide enough evidence that either amplitude modulation or comprehensible speech is responsible for infants' speech perception difficulties; additionally, the preliminary results did not replicate the existing finding that amplitude modulation and comprehensible speech in the background are more difficult for infants than speech-shaped noise, which provides energetic masking of the target speech but no informational masking.

It is possible that in order to see an accurate result for this study that I would need to collect approximately 140 infants, according to a power analysis of a similar study (Newman, 2009). However, none of the published studies using these methods have ever used close to this number of participants. If I collected 140 infants, I would have expected to see a main effect of Name, such that infants prefer to listen to their name overall more than the foil, regardless of background noise type. I would have potentially expected to see a main effect of Background Noise Type, as infants may listen longer overall to trials in which the background speech is more speech-like. I would have expected to see an interaction between Name and Background Noise Type, such that infants show a greater preference for their name in comparison to the foil in SSN as compared to both ST and AmpMod-SSN. I would have expected to see that infants prefer their name less in comparison to the foil name in Amp-Mod SSN and ST than SSN, because the presence of amplitude modulation and/or speech content should make these two conditions more distracting. The crucial finding in this study would be whether infants show a greater preference for their name in comparison to the foil in AmpMod-SSN in comparison to ST. Amp-Mod SSN contains the temporal fluctuations of ST without the speech content. If infants did not show a significant difference in name preference between these two noise conditions, but still showed a smaller name preference for both than that seen in SSN, it would indicate that infants listening to speech in the presence of a background talker are primarily distracted by the amplitude modulations, and that speech content does not add to their distraction. If infants preferred their name to the foil significantly more in the presence of AmpMod-SSN than ST (but a smaller name preference than that seen in SSN), it would have indicated that both amplitude modulation and the presence of comprehensible speech play a role in infants' difficulties perceiving speech in noise. If there was a large difference in infants' preferences between AmpMod-SSN and ST, it would have indicated that comprehensible speech is a large contributor to infants' difficulties. If there was a small difference between infants' performance in Amp-Mod-SSN and ST, it would have indicated that while comprehensible speech adds to infants' difficulties in noise, it is not the main contributor behind infants' poor performance.

However, it is possible that even after collecting 140 infants, there would be no significant result due to the study design. In this study, infants hear their own name much more often than in other similar studies (e.g., Newman, 2005; 2009). This type of experiment relies on infants being interested in listening to their own name; however, if they hear it too often, it might become less interesting. Other studies such as Newman (2005) and Newman (2009) utilize a design in which the infant hears four blocks of four trials consisting of their own name and three foils, all in the presence of background noise. In Newman's design, the infant only hears their name in four trials during the study. In the design used in this study, the infant heard their name nine times in total, albeit with different background noise. This may be too many repetitions to hold the infants' interest, despite the changing background noise.

In addition to examining the effect of temporal fluctuation and the presence of comprehensible speech, future studies can look further into the different aspects of the comprehensible speech (e.g., lexical and phonological interference) and how these aspects impact infants' attention. For example, infants may be more distracted by the mere presence of an intelligible voice than the content of the speech itself. Identifying the elements of background noise that are primarily responsible for infants' distraction while listening to speech can lead to better guidelines for caretakers and daycares to reduce especially distracting noise.

# Chapter 5: The role of linguistic experience in the development of the consonant bias

## *Overview*

If someone is telling you a story about an animal they saw recently, a *dunkey*, would you assume they are referring to a *monkey* or *donkey*? Both *monkey* and *donkey* refer to animals, and both differ from *dunkey* by one sound. Despite the similarities between these potential animal names, adults do not treat these possibilities as equally likely. Instead, they are more likely to assume that a *dunkey* refers to a donkey, rather than a monkey (Cutler, Sebastian-Galles, Soler-Vilageliu, & Van Ooijen, 2000). They will more readily accept a mispronunciation and access the intended target word when the mispronounced word differs in vowel and retains consonantal information (as in dunkey-donkey) than when it keeps the same vowel but differs in the consonant (as in monkey-dunkey). This greater reliance on consonantal information, in both identifying and learning words, is known as the *consonant bias*.

The consonant bias is a reliable finding in adults across different language backgrounds and in many different tasks (Cutler, Sebastian-Galles, Soler-Vilageliu, & Van Ooijen, 2000, for Spanish and Dutch; van Ooijen, 1996, for English). Indeed, its consistency in adults has led researchers to theorize that the two major speech sound categories in language, consonants and vowels, serve different purposes for speech perception (Nespor, Peña, & Mehler, 2003). Vowels provide more information about prosody and speaker identity, while consonants play a large role in determining word identity.

Studies suggest that the consonant bias emerges over the course of development (Delle Luche et al., 2014; Højen & Nazzi, 2016; Nazzi, Floccia, Moquet, & Butler, 2009; Poltrock & Nazzi, 2015), with very young infants typically showing the opposite pattern (a vowel bias; Bouchon, Floccia, Fux, Adda-Decker, & Nazzi, 2015). It is not entirely clear what drives these developmental changes. It is thought that linguistic experience is a necessary and important factor, since children raised in different language environments show different developmental patterns (e.g., Nishibayashi & Nazzi, 2016; Højen & Nazzi, 2016). However, some have argued that auditory development (particularly in the area of temporal resolution) may also play a role (Poltrock & Nazzi, 2015). Since both are developing in tandem in typically-developing infants, it is difficult to assess the effect of linguistic experience alone, and determine how much exposure is required to support the emergence of the consonant bias in infants.

To attempt to unravel these different causal factors, the current study uses a dog model to examine whether dogs' linguistic experience and minimal lexicon are sufficient for consonant bias development. Other nonhuman animals, without regular exposure to human language, typically show a vowel bias, akin to that initially shown by young infants (Bouchon & Toro, 2019). Bouchon et al. (2019) discussed the possibility that "given other appropriate experience and exposure to speech, a consonant bias would emerge in non-human animals" (p. 848). The adult domestic dog, a pet with a mature auditory system and long-term, natural exposure to language, provides an appropriate and useful model to assess whether regular exposure to language in addition to a limited lexicon is enough to spur the development of the consonant bias. We examined dogs' differential use of two categories of sounds in language, consonants and vowels, to determine word identity. By characterizing dogs' word representations, we can better understand what aspects of language can be learned through exposure to language via shared auditory processing capabilities, and what patterns can be uniquely attributed to the human linguistic system. Learning how dogs represent and perceive words can also better inform training practices for working dogs as well as companion animals.

### *The consonant bias in human infants*

While the consonant bias is a reliable finding in adult humans, it is unclear at what point the consonant bias emerges in development. Very young infants tend to show a vowel bias (Bouchon et al. 2015), suggesting that the consonant bias emerges either with experience or maturation. The vowel bias makes sense logically as a starting point: vowels are typically longer and louder than consonants (Ladefoged, 2001), thus are more acoustically salient for infants (Mehler, Dupoux, Nazzi, & Dehaene-Lambertz, 1996). An alteration to a vowel should then be more noticeable perceptually for infants, all other things being equal. However, a shift from a vowel to a consonant bias appears to occur relatively early in development. In word recognition tasks, studies have found a consonant bias in French infants as young as 11 months of age (Poltrock & Nazzi, 2015). Additionally, Nazzi (2005) found that 20-month-old French toddlers will assign two object labels that only differ in vowel to the same object, but will assign two objects labels that only differ in consonant to different objects, indicating that they are treating consonant differences, but not vowel differences, as crucial to word identity.

Despite the consistent pattern seen in French infants, the emergence of the consonant bias varies cross-linguistically. This suggests that the consonant bias is modulated by native language exposure. English-learning infants only reliably show a consonant bias at 30 months of age (Delle Luche et al., 2014), and there are inconsistent results in word recognition tasks, with some suggesting that English infants demonstrate a consonant bias at 14 months (Ballem & Plunkett, 2005), while others show that 15-month-olds do not (yet) demonstrate this bias (Mani & Plunkett, 2007). It is possible that this difference between French and English infants is due to the wide variety of accents present in Great Britain (where all prior studies of the consonant bias in English-hearing infants were conducted). As such, infants may have been tested with a voice that had an unfamiliar accent, which would make it harder for the infants to comprehend the speech they heard. As a result, it is unclear whether younger infants are not demonstrating the consonant bias due to difficulty comprehending the stimuli, or whether they genuinely do not develop this bias until they are much older than the French infants.

Danish infants are a particularly unique case with regards to this cross-linguistic variation. French and English both have many more consonants than vowels, resulting in these languages having more minimal pairs that must be distinguished by their consonants than ones that must be distinguished by their vowels (Hochmann, Benavides-Varela, Nespor, & Mehler, 2011). (Minimal pairs are pairs of words that differ by only a single phoneme; thus *cat* and *mat* are minimal pairs that differ in their consonants, while *cat* and *cut* are minimal pairs that differ in their vowels). Danish as a language features more vowels than consonants, and the consonants are often underarticulated, which further increases the utility of vowels for word identification. This may be why Danish infants fail to demonstrate a consonant bias entirely and instead demonstrate a vowel bias at 20 months in a word-learning paradigm (Højen & Nazzi, 2016). However, it is not known whether Danish infants eventually switch to a consonant bias or persist with a vowel bias; there are currently no studies that test for a consonant bias in older Danish children or Danish adults.

In addition to native language exposure, theories have suggested that the development of the auditory system also aids in the emergence of the consonant bias (Poltrock & Nazzi, 2015). One theory behind the emergence of the consonant bias in infants, the acoustic-phonetic hypothesis (Bouchon et al., 2015), suggests that the consonant bias emerges in infants due to exposure to both the different acoustic and phonetic properties of vowels and consonants. First, the consonant bias may begin to emerge due to the development of better temporal resolution in the auditory system, which allows for better perception of (often quickly-changing) consonantal information. Second, the consonant bias emergence is accelerated by the acquisition of native phonemic categories, which better indicate to the infants what consonants are informative in their language. Thus, this theory suggests that the emergence of the consonant bias may be driven by both auditory and linguistic development.

Another theory of consonant bias development, the lexical hypothesis (Keidel, Jenison, Kluender, & Seidenberg, 2007) focuses on the structure of the acquired lexicon. It suggests that as infants learn more words, the distributional information they learn about the words highlights the importance of consonantal information for word identity. This would then lead to privileged processing of consonants in comparison to vowels in languages with more consonantal minimal pairs.

One way to test these theories experimentally would be to hold auditory development constant while providing different amounts of linguistic experience to different individuals. For obvious ethical reasons, this approach cannot be taken with young infants. However, it can be implemented in non-human animals, who have mature auditory processing capabilities, can gain language exposure naturalistically or in experimentally controlled conditions, and can be taught word forms. In this fashion, we can evaluate competing theories concerning the effect of linguistic experience as well as the size and structure of the lexicon on consonant bias development.

### *Rats as a model for the consonant bias*

Prior work has examined whether rats would show privileged processing of consonantal information (Bouchon & Toro, 2019). Laboratory rats are a basic test case, as they have a mature auditory system but no linguistic system nor any long-term linguistic exposure to human speech. The authors argued that if rats showed a consonant bias, it would indicate that distinguishing between the physical and perceptual aspects of vowels and consonants alone allows listeners to determine that consonantal sounds are more useful for establishing word identity (Bouchon & Toro, 2019). Rats were trained to nose-poke a feeder when they heard trained word forms. Researchers then compared the number of times the rats nose-poked the feeder when presented with a trained word versus a novel word form. Two other item types were also tested: a consonant mispronunciation and a vowel mispronunciation of the trained word forms. The study concluded that rats demonstrated a vowel bias, in which the rats treated consonant mispronunciations more like familiar trained words than vowel mispronunciations (i.e., treating the consonant mispronunciation *pano* more similarly to the trained word *mano* than the vowel mispronunciation *mino*). This is a similar pattern to results seen in young infants, where it is interpreted as a vowel bias. Together, these studies show the importance of language exposure on the emergence of the consonant bias (Delle Luche et al., 2014; Højen & Nazzi, 2016; Nishibayashi & Nazzi, 2016; Poltrock & Nazzi, 2015).

The use of laboratory rats, who did not have any linguistic exposure prior to their word training sessions, only allowed for the conclusion that auditory processes alone are not sufficient for the consonant bias. This result is consistent with that of previous infant studies, which have demonstrated the importance of language exposure on the emergence of the consonant bias (Delle Luche et al., 2014; Højen & Nazzi, 2016; Nishibayashi & Nazzi, 2016; Poltrock & Nazzi, 2015). Given the results of Bouchon's study and prior infant studies, the consonant bias requires some degree of linguistic experience, but it is unclear to what degree more mature linguistic processing (in the form of phonological representations or a lexicon of a specific size and structure) is needed in comparison to mature auditory processing. It is possible that with the appropriate linguistic exposure, animals may develop a consonant bias (see Perez et al., 2013 for vowel and consonant differentiation in animals).

### *A domestic dog model of consonant bias emergence*

The domestic dog is a better animal model for testing the consonant bias, with two major advantages over the previous rat model. First, canine hearing is much more comparable to humans in their frequency discrimination and fine-grained temporal resolution (Bach et al., 2016), suggesting that they would be sensitive to many of the same cues for consonants and vowels as would young children. Second, dogs (at least in the US) are typically kept as pets within a human household, where they are naturally exposed to language input. This occurs both ambiently (from humans talking to one another in their environment) and from speech directed towards them (Burnham, Kitamura, & Vollmer-Conna, 2002). To test whether linguistic exposure is enough for the emergence of a consonant bias, it is necessary to select a model organism that receives persistent human language input over a long period of time, and can learn words from that language. The domestic dog is an ideal choice. Testing dogs allows for an examination of the contribution of linguistic experience and lexicon size and structure while controlling for auditory development.

The domestic dog has been an important model species for comparative work in recent years, including in studies of human speech perception (Andics et al., 2014; Mallikarjun, Shroads, & Newman, 2019). Through domestication, they have been selected across thousands of years to be attentive to human communicative behaviors (Hare et al., 2002); these include gaze, pointing gestures, and speech (Hare, Call, & Tomasello, 2010; Horowitz, 2009; Miklösi et al., 1998).

Dogs are not only exposed to and attend to language in their environment, but they also learn individual words (Griebel & Oller, 2012; Kaminski et al., 2004; Pilley & Reid, 2011). Some dogs may even acquire vocabularies that are similar in size to those of young children (Pilley & Reid, 2011). However, even dogs without special linguistic training have been shown to learn a number of different words. Pet dogs can recognize several commands, even at a young age (Kutsumi et al., 2012). Some of the words in a pet dog's lexicon are taught directly to the dog, like commands, and some the dog picks up via association (i.e., the dog learns that *walk* means they will go outside, because that is what usually happens when the owner says *walk*). Thus, if possessing a lexicon is a prerequisite for shifting to a consonant bias (Keidel et al., 2007), pet dogs may show this bias.

Moreover, pet dogs have been shown to learn properties of their most-often-heard, or "native", language. Studies from our lab have shown that dogs can differentiate their "native" language from unfamiliar languages that differ in rhythm and phonology, indicating that they have some awareness of the underlying features of their "native" language. This, too, suggests that dogs may have the linguistic exposure necessary to demonstrate a consonant bias.

Given dogs' mature auditory abilities in conjunction with their linguistic exposure, testing dogs' detection of consonant and vowel mispronunciations can help determine whether a smaller amount of linguistic exposure suffices for the emergence of the consonant bias. Conveniently, dogs can be tested using an identical method to one used to evaluate the consonant bias in infants, the Headturn Preference Procedure (HPP). HPP is an experimental paradigm generally used to test infants on their preferences for different auditory signals. In one study of the consonant bias in young infants using HPP as a method, infants' preferences were compared across three types of stimuli: their own name, a version of their name with the initial consonant in the stressed syllable mispronounced, and a version of their name with the vowel in the stressed syllable mispronounced (Bouchon et al., 2015). HPP has been used to demonstrate dogs' recognition of word forms (Mallikarjun et al., 2019); dogs were presented with their name or another dog's name as spoken by an unknown voice, and showed longer listening to their own name. In this study, dogs were presented with their own name or a mispronounced version, akin to the stimuli in Bouchon et al. (2015). This allows for an evaluation of whether dogs, with their linguistic exposure and limited lexicon, show a consonant bias, like adult humans and toddlers, or a vowel bias, like both young infants and rats.

### *Experiment 1: Dogs' listening times for their name with a vowel or consonant mispronunciation*

This study tests dogs' preference for their own name over their name with a mispronounced vowel or consonant in the initial (stressed) syllable. French infants can detect vowel mispronunciations in their name several months before they can detect consonant mispronunciations (Bouchon et al., 2015). Researchers suggest this is because vowels are more salient than consonants: they are louder, longer, continuous, sonorant, and more periodic in structure (Cutler & Mehler, 1993). As such, young infants may primarily focus on acoustic salience to differentiate word forms. Although similar studies have not been done with young infants in other language backgrounds, the presumption is that this early focus on acoustic salience would be universal across infants from all backgrounds. That is, young infants' low linguistic exposure, lack of native phonological categories, and poor temporal auditory processing skills would lead to infants of all language backgrounds to initially demonstrate a vowel bias; only with sufficient exposure to input prioritizing consonantal information would children's processing shift towards a consonant bias.

We expect that dogs will generally prefer to listen to their name over a mispronounced version, but this may vary depending on the type of mispronunciation. If dogs have a consonant bias, we would expect to see an interaction in which they show a stronger preference for their name in comparison to the version with a mispronunciation on the consonant than in comparison to the vowel mispronunciation. If instead, like the 5-month-olds in Bouchon's study, dogs primarily rely on acoustic salience to distinguish between words and have not developed a consonant bias, we would expect that they have a stronger preference for their name in comparison to the vowel mispronunciation rather than in comparison to the consonant mispronunciation.

**Participants.**

Forty-four dogs (23 M) participated. To be included in the study, dogs must have had their name for at least ten months prior to participating. We excluded any dogs that were taking psychiatric medication, and dogs whose owners noticed signs of hearing loss. On average, the dogs were 5.1 years old, and had been hearing their name for 4.8 years. Twelve of these dogs were therapy dogs, five were search-and-rescue dogs, and two were service-dogs-in-training. Only dogs with one-syllable or two-syllable trochaic (stressed-unstressed) names were included in this study. Thus, mispronunciations always occurred in an initial, stressed syllable.

Twenty-two dogs (11 M) participated in the Vowel Mispronunciation condition of this experiment, and twenty-two dogs (12 M) participated in the Consonant Mispronunciation condition of this experiment. Three additional dogs were tested but were excluded from the study: One due to experimenter error, and two due to noncompliance (e.g., failure to orient to sounds, falling asleep).

**Test materials.**

Prior to the study visit, each dog owner was asked the name or nickname that their dog is most commonly called; this was used as the dog's name in the study. Every dog heard four different trial types: his or her name, a mispronounced version of his or her name, a foil name that shares minimal phonetic characteristics with his or her name, and a mispronounced version of the foil name. Including a mispronounced version of the foil name ensures that, regardless of whether dogs notice a phonetic difference or not, there are equivalent numbers of trials that are familiar to the dog (i.e., dog's name, and potentially the mispronounced name) in comparison to trials of any given name that is perceived as novel. (That is, if dogs in the Vowel Mispronunciation condition ignore vowel differences, they hear half of the trials with their name and half without; if they do not ignore these differences, they hear ¼ of the trials with their name and ¼ of the trials with each of the other three names.)

Twenty-two of the dogs heard a mispronounced name in which two to three features of the vowel in the stressed syllable was changed. Tense/lax features were maintained, and height and frontness were always changed. However, English correlates roundness with frontness/backness, so rounding was changed when necessary to maintain natural English phonemic categories (Table 5).

In the other condition, twenty-one dogs heard a mispronounced name in which two features of the onset consonant was changed. The mispronounced consonant version of the name kept manner the same, changed place, and changed voicing (Table 6). By mistake, one dog heard a mispronounced name in which only one feature, voicing, was changed. Below, we run the analyses with and without this dog, and it does not change the results of our study.

Other than the single dog with one feature change, only dogs with names that began with a stop, fricative, or affricate participated in this study, so that it would be possible to always change place and voicing. Dogs with names beginning with a nasal or approximant did not participate in this study.

There were a few differences between the stimuli for this study and the infant name stimuli from Bouchon et al. (2015). First, instead of changing the first phoneme of the participants' name, we changed the vowel or consonant in the first syllable of the dog's name; this is because there were few dogs visiting the lab who had a vowel as the first phoneme of their name (approximately 8% of the total dogs that have visited since the inception of the lab). Secondly, we used a larger number of feature changes in this study than in the infant studies from Bouchon (2015) because we were initially unsure whether dogs would respond to a single-feature change in either vowels or consonants. For this reason, we wanted to change more features to ensure that the change would be salient for the dogs. As such, the number of features changed in this study were similar to the number of features changed in Bouchon's rat study (2019) rather than the infant study.

Table 5

*Vowel Mispronunciation*

| Original Vowel | Vowel Used in Mispronounced Name |
|---|---|
| i | o |
| eɪ | u |
| ε | ʊ |
| Æ | ʊ |
| u | eɪ |
| o | i |
| ɑ | eɪ |
| ɑi | ɑu |

Table 6

*Consonant Mispronunciation*

| Original Consonant | Consonant Used in Mispronounced Name |
|---|---|
| p | d |
| b | t |
| t | b |
| d | p |
| k | d |
| g | t |
| f | ð |
| v | Θ |
| s | ʒ |
| z | ʃ |
| dʒ | tʃ |

Four different female native English speakers produced recordings for this study. For each condition (vowel mispronunciation and consonant mispronunciation), the speakers recorded the names for five or six of the dogs. To minimize the possibility that speakers would unintentionally produce the dog's name in a more attractive manner than foil names, speakers were given names to record in sets, and were kept blind to which dog name(s) in each set would serve as a target name. Additionally, correctly-pronounced and mispronounced names were intermixed within each set of names, to ensure they were produced as similarly as possible. To obscure which names belonged to which category to the greatest extent possible, no mispronunciation was given in the same set as its corresponding correct pronunciation (to prevent speakers from attempting to guess which name was more likely). Because dog names in the US are highly diverse ("Most popular U.S. pet names", 2019), the names that were mispronounced and those that were not were likely less obvious to speakers than in analogous studies with infants. Recordings were made in a sound-attenuated room using a Shure SM51 microphone with a sample rate of 48kHz and bit depth of 32.

For each participant, one of the four speakers would record lively, dog-directed speech of the dog's name, mispronounced name, foil, and mispronounced foil. Each dog heard only one speaker produce all four of their trial types. A total of fifteen tokens were selected out of each of the original recordings. The name, mispronounced name, foil, and mispronounced foil tokens were chosen to match each other as closely as possible for pitch, duration, intonation contour, emotionality, and vocal quality, based on perceptual similarity. There was an initial silence of 0.5

seconds before the first name was spoken. Pauses between tokens of dog names were adjusted such that each file had the same overall duration of 22 seconds. Because pauses could vary in length based on the exact length of name tokens, and the overall amount of silence could vary slightly across files, matching for amplitude was performed by considering only the speech within the stream rather than the entire length of 22 seconds. Silent pauses were removed from a copy of the stream and the resulting file (containing only the speech) was adjusted to match a set average RMS amplitude; subsequently, the original stream containing pauses was amplified by the same amount. In this way, the speech within the name streams was always matched for average amplitude.

**Apparatus.**

The testing apparatus was identical to that described in Mallikarjun et al., 2019. The experiment took place in a six-foot by six-foot three-sided test booth with 4-foot-high walls made from pegboard. To ensure that the dog could not see the researchers over the booth, a curtain hung from the ceiling to the top of each of the booth walls. On the front wall of the booth, there was a hole for a camera. The camera recorded the testing sessions and allowed the coder to see the dog's behavior inside the booth via a computer monitor. In the center of the panel, above the camera, a light was mounted. The side walls each had a light mounted in the center and a speaker directly behind the light. These speakers played stimuli for the dog. A Mac computer was used by the researcher behind the front wall of the booth for coding. The researcher used a button box to start trials and code the dog's looking behavior.

**Procedure.**

The dogs and their owners were brought into the booth by an experimenter and the owner signed consent forms. The dogs sat on the owner's lap or directly in front of the owner, depending on their size and what made them the most comfortable. The dogs initially either sat facing towards the front of the booth (towards the camera) or facing the back of the booth (towards the owner). In either case, the dogs' attention was maintained as much as possible at a point equidistant from the two sides of the booth (where the loudspeakers were located). As a result, the dog's natural inclination upon hearing a sound through one of the two loudspeakers was to turn their head or body 90 degrees to face the source of sound. There were two practice trials, one from each of the two speakers on the sides of the booth, to familiarize the dogs with the procedure. It is common to use more than two practice trials in this paradigm for infant studies (e.g., Newman, 2005, 2009), but dogs can become easily distracted and lose interest quickly with more practice trials, so only two were used here. Dogs' listening time was judged by the amount of time they spent looking at the sound source (the wall behind which the speaker was mounted). The practice trials featured a happy, friendly female voice talking to and praising the dog. This voice was never used as a target voice in the test trials.

The test phase began immediately after the practice trials. Dogs heard four types of stimuli: repetitions of their own name, a foil name, their name with a mispronunciation, and the foil name with a mispronunciation. Each stimulus type was heard on four separate trials for a total of sixteen trials, presented in four, four-trial blocks (one of each type of trial per block). Order of trials within each block was randomized. Two experimenters ran the test phase portion of the study: one to code the dog's looks (the coder), and one to produce auditory attention getters (the attention experimenter). The auditory attention-getters consisted of scratching noises, knocking, whistling, and squeaky dog toy sounds.

At the start of each test trial, the light on the front of the booth turned on, and the attention experimenter made a sound to get the dog's attention to the front of the booth. Although work with infants typically uses only lights as attention-getters, pilot work suggested that the light alone was not sufficient for most dogs. The light also served as the apparent "source" of the sound for the dog, and helped the coder code the dogs' looks to the sound source. Once the dog attended to the front, that light turned off, the light on either the left or right side of the booth turned on, and the attention experimenter made a sound on that side. Once the dog attended to that side, the stimulus for that trial began to play from the loudspeaker on that side. The coder used a button box to code the dog's looks towards and away from the sides. The stimulus continued to play for a full twenty-two seconds or until the dog looked away for two consecutive seconds, whichever occurred first. Any time the dog spent looking away was subtracted from the dog's overall looking time. The coder wore Peltor aviation headphones playing masking music so she would not be able to hear the trials and have that influence her coding.

**Results and preliminary discussion.**

A 2 x 3 mixed ANOVA was used to test the effect of Condition (Vowel, Consonant) and Item (Name, Mispronounced Name, Foil) on listening time. The Foil and Mispronounced Foil were combined into a single category (Foil), as they were both equally unfamiliar to the dog. We found no main effect of Condition, $F(1, 43) = 1.05$, p. > .05, or Item, $F(2, 43) = 1.06$, $p > .05$, but did find a significant interaction between Condition and Item, $F(2, 43) = 4.099$, $p < .05$. To determine the nature of this interaction, individual 1 x 3 within-subjects ANOVAs were conducted in each condition.

For the dogs in the Vowel condition, a 1 x 3 within-subjects ANOVA was used to test the effect of Item (Name, Mispronounced Name, Foil) on listening time. We found an overall effect of Item, $F(1, 21) = 4$, $p < .05$. Dogs listened longer to Name trials (7.37 seconds) than the Mispronounced Name (5.57 seconds; $t(21)=2.66$, $p < .05$) or Foils (5.91 seconds; $t(21) = 2.28$, $p < .05$). There was no significant difference in looking time between the Mispronounced Name and Foils, $t(21) = -.611$, $p > .05$. Additionally, the foil and the mispronounced foil, which were averaged together in this analysis, did not differ from each other ($t(21) = -.008$, $p > .05$). Figure 16, below, shows a graph of these results. Like young infants and rats, dogs are treating a change in vowel as though it changes the meaning of a word.

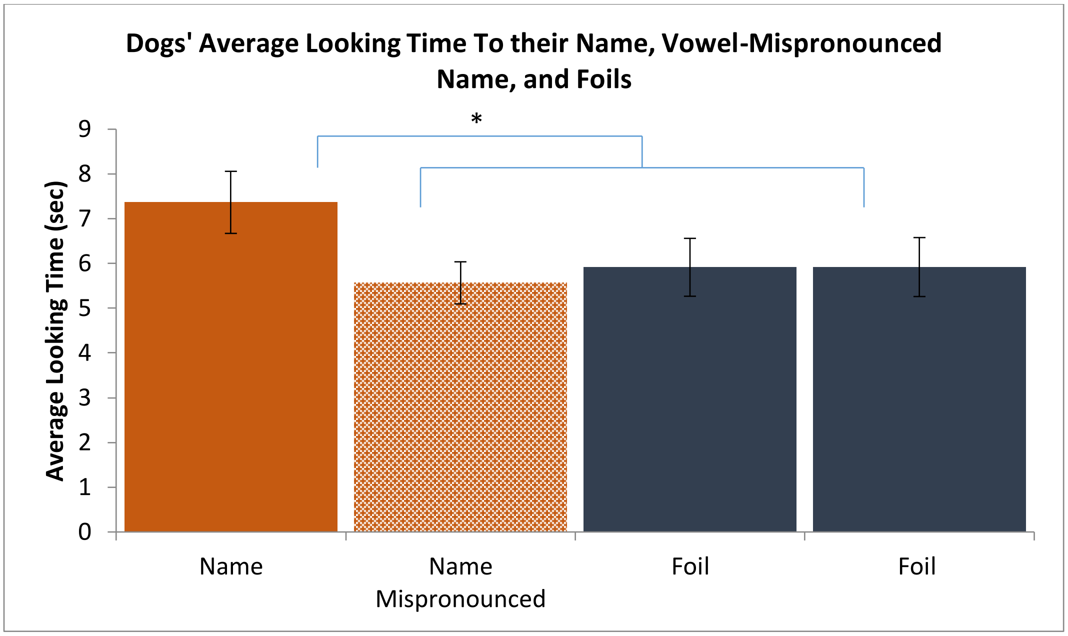


Figure 16: A graph of dogs' average looking time in seconds to their name, their name with a vowel mispronunciation, and two foils (one foil was a mispronounced version of the other foil). Dogs preferred to listen to their name rather than the mispronounced name or foils.

For the dogs in the Consonant condition, a 1x3 within-subjects ANOVA was used to test the effect of Item (Name, Mispronounced Name, Foil) on listening time (see Figure 17, below, for a graph of the results). We found no effect of Item, $F(1, 21) = 0.717$, $p > .05$. This suggests that dogs may not notice a change in consonant, much like young infants (Bouchon et al., 2015) and rats (Bouchon & Toro, 2019).

There was a single dog in the Consonant condition with only one feature change to create the mispronounced name rather than two feature changes. When this dog is removed, the results of the analysis do not change (no effect of Item, $F(1, 20) = 1.45$, $p > .05$).

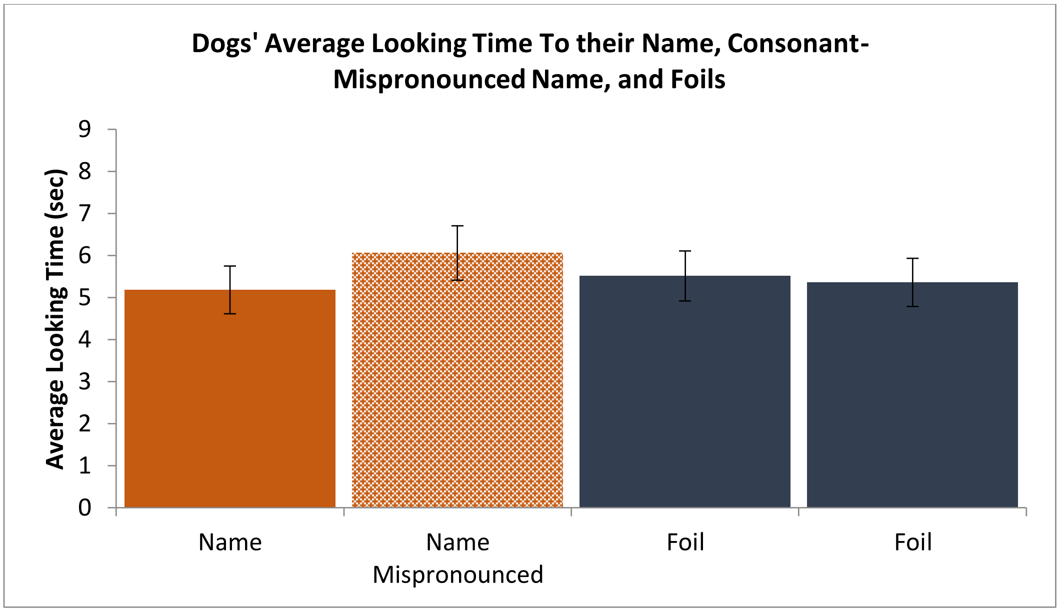


Figure 17: A graph of dogs' average looking time in seconds to their name, their name with a consonant mispronunciation, and two foils (one foil was a mispronounced version of the other foil). Overall, there was no effect of Item (Name, Name Mispronounced, Foils).

However, there is one aspect of these Consonant condition results that is surprising: not only did dogs not prefer their name over the version with a consonant mispronunciation, they also did not prefer their name over the foil, which clearly differed in many ways from their own name.

One possibility is that dogs may consider the mispronunciation to be the actual equivalent of their name. That is, the dogs may not be able to perceive the mispronunciations. If so, they would hear only two trial types in this study, Name trials and Foil trials, as opposed to the four different types of trials, Name, Foil, Name Mispronounced, and Foil Mispronounced, that we anticipated they would perceive. This could lead them to get bored much more quickly in the study; instead of hearing each of four trial types four times, they perceived each of the two trial types eight times each.

If this were the case, one might expect that the first two blocks (the first 4 repetitions of each of the two perceived names) would show an effect, even if the full experiment did not. (That is, since dogs in the vowel condition showed a name vs. foil preference with 4 repetitions of each item, as did dogs in Mallikarjun et al. (2019), we might expect the dogs in the current study to do likewise.) We therefore examined dogs' preference in just the first two blocks of this experiment to see if they showed the basic preference for their name over a foil name (see Figure 18). We did this analysis two ways: first, using the same 1x3 within-subjects ANOVA we used before (Name, Mispronounced Name, Foil), and second, collapsing name & mispronounced name, and comparing this to the combination of foil and mispronounced foil. We found no effect in either case (1x3 ANOVA: $F(2, 41) = 0.523$, $p > .05$; t-test: $t(41) = 1.05$, $p > .05$). Thus, even in the first two blocks, dogs in this study did not show a preference for their name over the foil name. It is not clear what to make of this pattern; it might suggest that dogs do not necessarily consider the name and mispronounced name or the foil and mispronounced foil to be equivalent. If the lack of preference for Name trials over Foil trials was just due to boredom because of perceived repetition, we would expect to see the Name preference in the earlier trials. Instead, we see no effect at all. Interestingly, we have seen this same pattern in other canine studies; for example, in Mallikarjun et al., 2019, dogs heard their name and a foil name in quiet and their name and a foil name in the presence of background noise. When the noise level was low, dogs showed a preference for their name over the foil in *both* quiet and noise conditions. When the noise level was more intense, dogs not only stopped showing the preference for their name in noise, but also in quiet. Thus, it appears to be a relatively consistent finding that when a task becomes very difficult, dogs appear to "give up" on the study and look for short, equal periods towards all trials (Mallikarjun et al., 2019). As such, the failure to show a preference for name over foil may be an indication that the inclusion of items that differ only by a consonant makes the task itself more difficult.

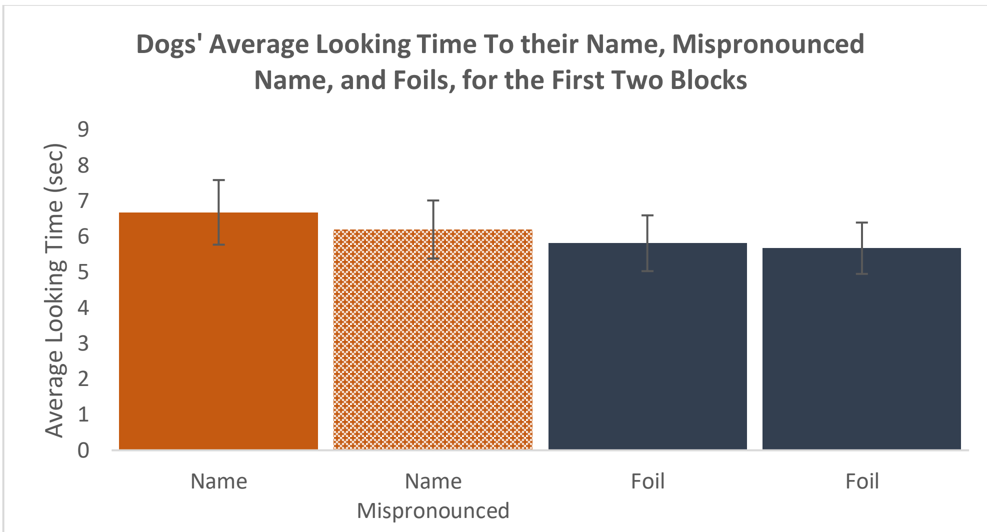


Figure 18: A graph of dogs' average looking time in seconds for the first two blocks of the consonant condition to their name, their name with a consonant mispronunciation, and two foils (one foil was a mispronounced version of the other foil). Overall, there was no effect of Item (Name, Name Mispronounced, Foils).

It is worth noting that there were two to three featural changes in the vowel items (Height, Front-Back, and sometimes Rounding) and only two in the consonant items (Place and Voicing). It is possible that the fewer feature changes in the consonant condition would make the consonant condition harder than the vowel condition. However, consonant categories are generally more acoustically distinct from each other than vowel categories. A spectral analysis with vowels and consonants done by Bouchon et al. (2015) indicated that, when normalized for duration and intensity, two contrasting consonants that differ in a single feature are more acoustically distinct than two contrasting vowels with a single feature change, meaning that the two consonants are easier to distinguish from one another than the vowels. As a result, even with an additional feature change in the vowel mispronunciations, it is not necessarily the case that the consonants would be less acoustically distinct than the vowels.

Thus, the current results suggest that while dogs notice the difference between their name and one with a vowel change, they have more difficulty doing so when the names differ only in a consonant. Experiment 2 seeks to explore this issue more deeply, by examining whether dogs treat an item with a consonant mispronunciation as if it were their own name, in cases where discrimination among items is easier.

### *Experiment 2: Preference for a Name with a Consonant Mispronunciation in the Absence of the Correctly Pronounced Name*

This study uses a different approach to determine whether dogs detect consonant mispronunciations in their name. The prior experiment suggests that while dogs prefer their name to one that has a vowel mispronunciation, they do not show a preference for their name compared to one with a consonant mispronunciation. This might suggest that the item with only a consonant mispronunciation is close enough to "count" as their name. In the current experiment, dogs are presented with the mispronounced version of their name and three foils; they never hear a correctly-produced version of their name. If dogs consider their consonant-mispronounced name to be more similar to their actual name than the foil names, we would expect them to listen longer to the mispronounced version of their name than the foils.

**Participants.**

Twenty-two dogs (11 M) were tested in this study. We excluded any dogs that were taking psychiatric medication, and dogs whose owners noticed signs of hearing loss. On average, the dogs were 4.4 years old, and had heard their name for 4.2 years. Only dogs with one-syllable or two-syllable trochaic (stressed-unstressed) names were included in this study. Thus, mispronunciations always occurred in an initial, stressed syllable.

Three of these dogs were therapy dogs. Six dogs were excluded due to owner interference in the study (1), equipment error (1), and noncompliance during the study (4).

**Test materials.**

The consonant-mispronounced version of the dog's name was created in the same manner as the consonant mispronunciation version of Experiment 1.

Unlike Experiment 1, this experiment does not utilize the dog's actual name. The stimuli the dogs heard consisted of the mispronounced version of the dog's own name, as well as three other dogs' names or mispronounced names that served as foils. The foils were selected to maximize perceptual dissimilarity between the consonants and vowels in the mispronounced name and the foil names. As such, 11 participants heard exclusively mispronounced foil names in addition to their own mispronounced name, and 11 participants heard a combination of correctly pronounced and mispronounced dog names in addition to their own mispronounced name; we assumed that dogs would not know whether other names were "standard" vs. mispronounced. (Since we do not tell our speakers which names are mispronounced and which are not, we do not anticipate that dogs will listen any longer to “real” names than mispronounced names.) The names were recorded and edited in the same manner as Experiment 1.

**Apparatus.**

This study was run using the same method as the previous study (HPP) but the testing apparatus was moved to a different room and the software was updated (Newman et al., 2019). The setup remained almost identical, with three small changes: in this study, a GoPro was used instead of a low-light security camera to record the testing sessions, and a Windows computer instead of a Mac was used for coding. A keyboard, rather than a button box, was used to code the dogs’ looking behavior.

**Procedure**

Same as Experiment 1.

**Results**

A 1x2 within-subjects ANOVA was used to test the effect of Item (Mispronounced Name, Foil) on listening time. Figure 19, below, shows a graph of the results. A main effect of Item was found, $F(1, 21) = 6.01$, $p < .05$, where dogs look longer at the Mispronounced Name (8.17 seconds) than the Foils (an average of 6.26 seconds over the three foils). Dogs prefer the mispronounced version of their own name to unfamiliar, phonetically dissimilar foil names. It is not possible with this study to distinguish whether dogs actually believe the mispronounced version of their name is their name, or can detect differences but decide to listen to the mispronounced name regardless. However, given that dogs have previously demonstrated the ability to distinguish between consonants that differ only in one feature (Adams et al., 1987), it is more likely that they can distinguish between this mispronounced name and their true name, but do not consider this difference meaningful.

These findings support the notion that dogs show a vowel bias, as they preferentially attend to vowel information in determining word identity. While dogs in Experiment 1 preferred listening to their own name rather than a version with a vowel mispronunciation, they did not do so for consonant mispronunciations, where the vowel remained the same but the consonant changed. Moreover, the current findings suggest that they do not perceive a change in a consonant to be a critical difference that changes the meaning of a word (making a word no longer a match to the representation of their own name). This implies that changes in consonants are being treated as "less important" in some sense than are changes in vowels.

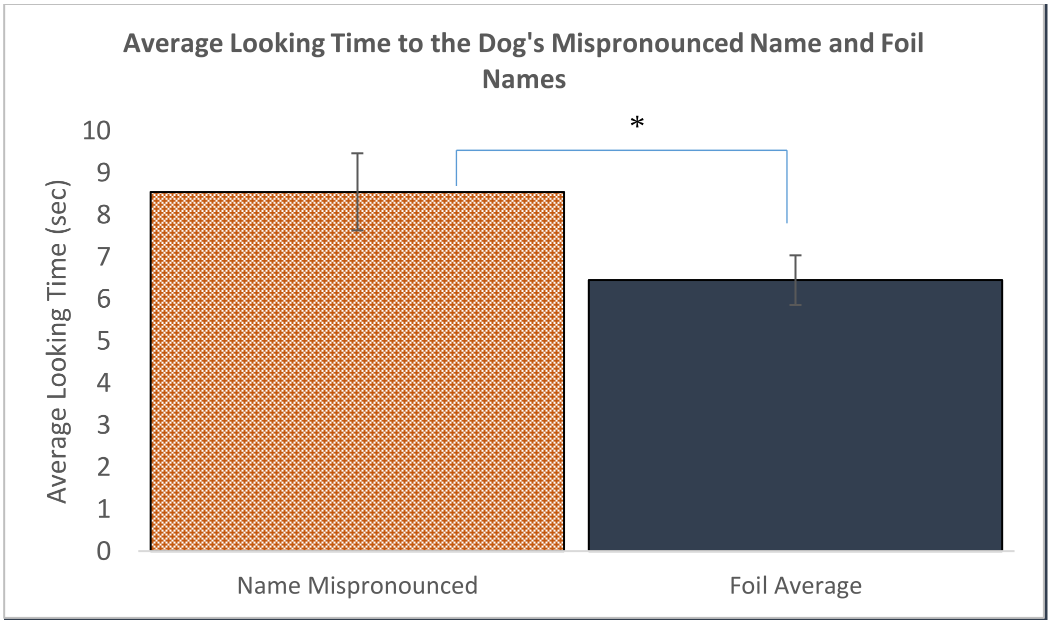


Figure 19: A graph of dogs' average looking time in seconds to the consonant-mispronounced version of their name and an average of their looking time to three foils. Overall, dogs listened longer to the mispronounced version of their name than the foils.

### *Overall discussion*

The goal of this study was to assess in a domestic dog model whether linguistic experience and a small lexicon suffices for the consonant bias to emerge. The results indicated that despite their linguistic experience, dogs did not demonstrate a consonant bias; they treated a version of their name with the initial consonant changed as essentially equivalent to their actual name. Instead, dogs showed a vowel bias, as they distinguished between their actual name and a version of their name with a change to the vowel in the stressed syllable. This is the same result seen in rats; Bouchon et al. (2015) found that rats similarly showed a vowel bias rather than a consonant bias. They argued that mature auditory processing in the absence of a lexicon or consistent linguistic exposure was not enough for consonant bias development. However, even with additional linguistic exposure, dogs fail to show a consonant bias.

While young infants also demonstrate a vowel bias, like rats and dogs, infants generally switch from a vowel bias to a consonant bias between 8 and 15 months, depending on their native language (Nishibayashi & Nazzi, 2016; Poltrock & Nazzi, 2015). Why do most infants eventually develop this consonant bias when dogs do not? During this time period of development, infants gain more language exposure, learn more word forms, develop their native speech sound categories, and improve their auditory processing abilities. While dogs' auditory processing abilities are mature, they likely differ from infants in the nature of the language exposure, the types of words that they learn, and in the process of developing speech sound categories.

While dogs, like infants, are often in an environment with a great deal of linguistic input, they are unlikely to listen to and process the input in the same way as infants. Dogs are certainly interested in human language, and have been shown to have specific brain regions for processing the emotional valence and words of human speech (Andics et al., 2016, 2014). However, dogs' attunement to human communication is unlikely to rival that of human infants. Moreover, dogs likely do not receive as much direct speech as infants (although the exact amount of speech directed to dogs on a daily basis is unknown) and may be less attuned to overheard speech. Recordings of speech directed to dogs in their homes and throughout their day will be needed to assess these questions.

Another potential reason why dogs might not develop a consonant bias is that pet dogs do not have a large enough or varied enough vocabulary. The lexical hypothesis suggests that knowledge of more words and different kinds of words leads to increased distributional information about consonants and vowels within the words (Bonatti, Peña, Nespor, & Mehler, 2005); this in turn will help learners notice the importance of consonants for word identity. While most pet dogs know some words, they may not have a large enough vocabulary to trigger this reorganization. One way to explore this would be to look at the consonant bias in particular dogs that have been trained to have exceptionally large vocabularies. Indeed, there is some anecdotal evidence to suggest that dogs with large vocabularies may be more sensitive to changes in consonants. While the pet dogs in the current study failed to differentiate between their name and a version with a salient consonant change, Chaser, a border collie who knew over 1000 words, was able to differentiate between toys whose names were consonant minimal pairs (e.g., *tote* and *goat*) (Pilley & Reid, 2011). Chaser also differentiated between toys with vowel minimal pairs (e.g., *boo* and *bow*), suggesting this may represent a greater sensitivity to phonetic differences among consonants rather than a shift from a vowel focus to a consonant focus. Additionally, some of Chaser's known consonant minimal pairs differed by only a single feature, and some even had minimal pairs in non-stressed syllables (e.g., *odie* and *obie*). Thus, for at least one dog, a large vocabulary seems to correlate with more successful detection and learning of meaningful consonant changes within words.

It is possible that Chaser was able to learn this large number of object names because she could successfully treat these consonant changes as important for word identity. But it is also possible that the pressure to learn the words led her to gain this ability. Thiessen and Yee (2010) suggest that experiencing phonemes in several different lexical contexts allows infants to better understand and notice the relevant phonemic contrasts. For example, hearing the d/t contrast in *duck* and *tummy* in the same vowel context might help infants detect the distinction between the minimal-pair words *bun* and *done.* Chaser's vocabulary contained a wide variety of sounds in different contexts. Her known words ranged from one to six syllables long, with varying stress patterns (e.g., *firecracker* and *gingerbreadman*). This variety of contexts may have helped Chaser learn to better differentiate consonant contrasts. Future studies could train dogs on certain contrasts by presenting them with phonemes in different lexical contexts and seeing whether that experience helps them differentiate that contrast in a novel context.

Another potential reason dogs do not develop a consonant bias is that dogs may not be able to learn native phonological categories for sounds in their ambient language; this would make it more difficult to differentiate between consonant sounds, and harder to identify the role of consonants in determining word identity. The acoustic-phonetic hypothesis suggests that the development of native phonetic categories, which makes it easier for infants to categorize native-language consonant sounds, provokes the switch from a vowel bias to a consonant bias. This means it would be easier for infants to realize that /k/ sounds produced by two different people are in the same sound category, while a /k/ and a /g/, even if produced by the same person, are different sound categories. In learning native phonetic categories, infants must also ignore phonetic sound categories that may be meaningful in other languages, but are not meaningful in their own. For example, an infant learning Hindi must recognize the distinction between a dental /d/ sound and a retroflex /d/ sound. An infant learning English would not need to learn this distinction, and would assimilate these two sounds into the same category. It is unknown whether any non-human species can narrow phonetic discrimination of sounds after exposure to a language to form a native language inventory (Yip, 2006). Further experiments could assess whether dogs assimilate sounds that are not in their native language (given that English-hearing dogs distinguish between vowels like [a] and [i], would they not perceive a difference between the Danish vowels [i], like in *beet*, and [y], a rounded version of that sound that is not a unique phoneme in English?) This would help determine whether dogs narrow their phonetic perception after prolonged linguistic exposure.

These results also may have relevance for dog owners and trainers. Since dogs have more difficulty differentiating between words that differ only in a consonant, auditory commands given to dogs, especially those that may not appear with a visual signal, should differ from one another in their vowels, or dogs may have difficulty distinguishing between them. If commands do differ in consonant alone, like *bow* and *down*, an accompanying visual signal can aid in differentiation; the current results suggest that without such a visual cue, such commands may be difficult for dogs to learn. Similarly, when selecting names for dogs, it would be best if these names differed from common commands or from names of other household members in their vowels, for ease of differentiation (e.g., having a dog named “Pitt,” similar to *sit*, may be a poor choice, as would having two dogs named Rosie and Toby).

Beyond the consonant bias, further research can also explore whether experience with language shapes dogs’ detection and use of language-specific phonological and prosodic cues. For example, while English does not use tone as a contrastive feature that defines word identity, tonal languages like Mandarin and Thai do. Researchers could examine whether dogs raised in a household where they primarily hear a tonal language use tone to distinguish between words, and whether that is different from the way dogs raised in an English-speaking household treat tonal cues.

This study contributes to our understanding of the type of experience necessary for the emergence of a consonant bias in speech perception. Future studies will continue to explore the structure of speech input that allows for a consonant bias to emerge in infants, and whether, given similar input, non-human animals also develop a consonant bias. This will help to determine whether or not the consonant bias is a uniquely human phenomenon.

# Chapter 6: General discussion and conclusion

## *Overview*

This thesis set out to evaluate the utility of a comparative dog model to examine questions in the realm of developmental speech perception. This work was motivated by the problem that many questions in developmental speech perception are not possible to answer by testing exclusively infants, or through comparison with adults. The few animal models that have been used to examine developmental speech perception generally focus on whether certain speech perception abilities are human-specific or not. As such, this work was intended to assess domestic dogs as a potentially better comparative animal model for questions in developmental speech perception. I proposed that a dog model would be especially useful to assess 1) the contributions of individual underlying systems to different speech perception tasks; and 2) the role of language experience in the development of speech perception. Below, I will discuss empirical findings that address the contributions of underlying systems and the role of language experience in speech perception, and the implications of these results on the future use of dog models for these questions.

## *Research about systems underlying speech perception*

One goal of this thesis was to evaluate the use of a comparative dog model to identify how individual systems and processes contribute to infants' speech perception abilities. It can be difficult to tease apart the individual contributions of auditory, attentional, and linguistic processes to different speech perception tasks when testing only infants. In Chapters 2 and 3, I used a dog model to examine the individual effects of linguistic processing (Ch 2) and auditory versus attentional processing (Ch 3) on speech perception in different types of background noise. The results from Chapter 3 then allowed for a further study with infants (Chapter 4) examining the components of background noise that contribute to distraction and more effortful attentional processing. I will first provide a summary of the results from Chapters 2, 3, and 4. I will then discuss future studies motivated by the results of Chapters 2, 3, and 4, followed by a discussion of the implications the results have for future comparative dog model studies that examine the roles of underlying systems in developmental speech perception.

In Chapter 2, a domestic dog model is used to assess whether immature linguistic ability is the primary reason infants struggle to perceive speech in multitalker background noise. Dogs have an auditory system similar to that of adult humans, but do not have human-like linguistic processing; as such, if immature linguistic ability was a primary cause, it would be expected that dogs would perform similarly to infants. Dogs not only could recognize their name in multitalker background noise, but could do so at more difficult signal-to-noise ratios than infants. This demonstrates that limited linguistic abilities are not sufficient to explain infants' difficulties with speech perception in noise. Furthermore, dogs' ability to recognize their name at more difficult SNR levels than infants suggests that part of infants' difficulties with speech perception in noise stems from immature auditory processing. Chapter 2 additionally determines the signal-to-noise ratio (0 db SNR) at which dogs would have an equivalent auditory experience when listening to speech in the presence of background noise to that of 13-month-old infants (5 dB SNR) (Newman,

2005) and adults (-5 dB SNR) (Festen & Plomp, 1990). This is especially useful for researchers who want to compare dogs, infants, and/or adults, and construct similar listening experiences for each of these groups.

In Chapter 3, a domestic dog model was used to assess whether auditory or attentional immaturity is primarily responsible for infants' increased difficulty perceiving speech in single-talker background noise as compared to nine-talker background noise. Studies have suggested that infants' performance in single-talker background noise is especially poor due to immature stream segregation (a primarily auditory process) and immature attention. While dogs have a mature auditory system similar to that of adult humans, their auditory attention for human speech is closer to that of infants. Dogs performed more similarly to infants, rather than human adults, in their perception of speech in single-talker background noise; this result suggests that attention is the primary system responsible for infants' difficulties in single-talker background noise.

After the dog studies in Chapter 3 suggested that attentional factors were primarily responsible for infants' added difficulties with single-talker background noise, Chapter 4 examined the features of single-talker background noise that are most distracting for infants. This study did not have enough participants due to COVID-19, and did not show any significant results. It is possible that with further data collection this study may provide evidence to determine which aspects of single-talker background noise lead to the most distraction in infants. Importantly, this study shows how infant studies can build upon comparative work with dogs to answer questions that without the use of a comparative dog model would be difficult to assess.

**Future directions.**

One immediate future direction for this work is to complete and expand upon the study done in Chapter 4: studying the contribution of different spectral and temporal aspects of speech and background to infants' difficulties comprehending speech in noise. Chapter 3 demonstrated that infants' immature attentional processes are a large contributing factor to their additional difficulties understanding speech in single-talker background noise. However, it is unclear what aspects of single-talker background noise, as compared to nine-talker background noise, make it particularly attentionally challenging for infants. Chapter 4 attempted to examine the individual effects of spectral similarity between the noise and target speech, temporal fluctuation in the noise, and interference from comprehensible speech in the noise on target speech perception. The first future step would be to finish this study. However, given the way the study is designed, it would be hard to differentiate the effect of comprehensible speech in the background from the mere presence of an interesting, potentially distracting human voice in the background. An additional study could specifically test infants on their ability to recognize their name in the presence of an English-speaking single talker and a single talker speaking a different language. In this way, both conditions feature an interesting voice, but only one has potentially comprehensible speech and recognizable phonology for the infant. Together, these studies would clarify what particular features of single-talker background noise make it especially challenging for infants.

Another future direction for this research would be to explore the effect of experience with background noise on speech-perception-in-noise performance. Studies have shown that infants and children have more difficulty understanding speech and learning words in the presence of background noise than adults (Finitzo-Hieber & Tillman, 1978; McMillan & Saffran, 2016; Newman, 2005, 2009; Riley & McGregor, 2012). However, there are no studies exploring whether infants' speech perception in noise varies based on their individual prior experience with noise. In adults, practice with speech perception in noise via training programs leads to better overall speech-perception-in-noise abilities (Bugannim, Roth, Zechoval, & Kishon-Rabin, 2019; Kuchinsky et al., 2014) and one study has shown that this improvement remains even six months post-training (Song, Skoe, Banai, & Kraus, 2012). These training tasks may not correlate to real-life experience with noise, however, and infants' experience may be different than that of adults. There are also many potential confounding factors when trying to assess the effect of infants' experience with background noise on their speech perception in noise; for example, background noise level is correlated with socioeconomic status (Evans, 2004; Evans & Kantrowitz, 2002; Haines, Stansfeld, Head, & Job, 2002; Lapierre, Piotrowski, & Linebarger, 2012). A dog model would be useful in this case, as researchers can track related dogs from puppyhood that have gone to louder and more quiet homes. Dogs' training can also be held constant; for example, a study could test young service-dogs-in-training living with different raisers. While these dogs have the same training schedule, their ambient noise exposure at home may be vastly different. Longitudinally testing dogs' speech-in-noise abilities as a function of their noise exposure can tell us if there is a benefit to prior noise exposure for speech perception in noise.

Lastly, future research should conduct foundational work to understand underlying dog and infant systems; this will allow researchers to more precisely know what kinds of comparative questions they can ask using a dog model. Chapters 2, 3, and 4 together demonstrate that a comparative dog model is useful for broad research questions that aim to disentangle auditory, attentional, and linguistic processes in specific speech perception tasks; however, the potential research questions are limited by gaps in the current foundational knowledge about both dogs and infants. Current knowledge about infant and dog auditory, attention, and linguistic processes suggests that broadly, dogs have more mature auditory processes than infants (Strain, 2012; Werner, 2007), dog and infant auditory attention is relatively similar (Mallikarjun et al., 2019), and infant linguistic abilities, while immature, quickly outpace those of dogs (see McMurray, 2007, for the onset of infant's rapid vocabulary growth; Pilley & Reid, 2011, for relatively slow, effortful dog vocabulary growth). (McMurray, 2007). These similarities and differences can allow for some useful studies, like Chapters 2 and 3, that identify the broad contributions of certain underlying systems to difficulties in infant speech perception. For example, one relevant question about infant speech perception in noise is whether infants, when listening to two simultaneous streams, segregate the streams and attend to both simultaneously, or segregate the streams and attenuate the less interesting stream in favor of focusing attention on the more interesting stream. This process would require both the auditory system (assigning sounds to the appropriate stream) and the attention system (attending to both streams, or attenuating one stream in favor of another). In order to use a dog model to separate these two potential attentional strategies, it would be necessary to better understand the dog attention system. While there are some basic similarities in the length of time infants and dogs attend to speech stimuli (see Newman, 2009, and Mallikarjun et al., 2019), dogs may listen to human speech for similar amount of time as infants because their attentional capacities are similar, or because they have a higher attentional capacity, but their interest for human speech is lower than that of human infants. As such, dogs might have a different attentional strategy than infants due to different underlying capabilities. This demonstrates the need for more studies that examine and compare basic speech perception functions in both dogs and infants such that future studies can more effectively use dog models to test more complex speech perception questions in infants.

### *Research about the role of language experience*

A second goal of this thesis was to evaluate the use of a comparative dog model to answer questions about the role of language experience in speech perception tasks. It is difficult to specifically assess the role of language experience in infants, as their increased experience occurs alongside their auditory and attentional development. As such, it is hard to determine whether they develop a particular linguistic ability due to increased maturity of the underlying systems, or whether language exposure and increased experience plays a role. In Chapter 5, a domestic dog model was used to assess the effect of linguistic experience on the development of the consonant bias. Despite dogs' regular linguistic exposure in addition to their small lexicon, dogs fail to show a consonant bias in their perception of mispronounced versions of their name. Rather, dogs show a vowel bias, which is the same pattern seen in young infants as well as rats (Bouchon & Toro, 2019). This confirms prior findings that mature auditory processes alone are not sufficient for the emergence of the consonant bias in infants.

More generally, chapter 5 demonstrates that when assessing the role of language experience in speech perception, dogs can be useful for examining 1) the effect of a smaller lexicon on certain aspects of speech perception; and 2) the effect of regular linguistic exposure in the absence of a human linguistic system on certain aspects of speech perception. Pet dogs have long-term exposure to human speech (Mitchell, 2001) and have been shown to learn words both through direct instruction (Griebel & Oller, 2012; Kaminski et al., 2004; Pilley & Reid, 2011) and from overheard speech (Fugazza & Miklósi, 2020). Pet dogs also generally have a smaller lexicon than human infants; this allows for researchers to design comparative studies examining the effect of a small lexicon and exposure to speech on speech perception using a dog model.

**Future directions.**

Immediate future studies can more specifically explore the role of lexicon size and structure in the emergence of the consonant bias. The lexical hypothesis suggests that infants may develop a consonant bias when the structure of their lexicon statistically demonstrates that words are better distinguished by consonants than vowels (*lexical hypothesis*, Keidel et al., 2007). As such, if the known words and word forms in dogs' lexicons can be just as easily distinguished using vowels, which are more salient, it would be unnecessary for them to pay attention to the less-salient consonants. To test this, we could first examine the lexicon structure of high-vocabulary dogs like Rico and Chaser to determine whether that structure predicts that consonants are more informative than vowels. If that structure suggests that consonants are more predictive than vowels, we could test these dogs in the same name mispronunciation study done in Chapter 5 to determine whether they demonstrate a consonant or vowel bias. If lexicon structure plays a large role in consonant bias emergence, we would expect that high-vocabulary dogs would show a consonant bias. However, if the consonant bias requires other skills, such as the acquisition of native phonemic categories (Floccia, Nazzi, Delle Luche, Poltrock, & Goslin, 2014) the high-vocabulary dogs may still demonstrate a vowel bias.

Beyond the consonant bias, further research can examine the role of language experience in the development of word form representations. In order to recognize a word, the listener's representation of the word must be broad enough to include variations in a speaker's pronunciation or differences between speakers, but narrow enough to exclude any phonological changes that alter the meaning. Past studies have suggested that infants initially store more detail in their word representations than adults would, which can cause a failure to generalize to other examples of the word form (for example, if infants store the word form *cup* with information about the specific speaker, they may fail to generalize when another speaker of a different gender produces *cup*, or if a speaker produces *cup* with a different affect) (Houston & Jusczyk, 2000; Singh et al., 2002). With more language experience, infants learn what aspects are necessary to store in their representations and which are not. There are several questions we can ask about word representations using a dog model. First, by testing whether dogs make some of the same errors as infants in generalizing newly learned word forms when spoken by a novel speaker or produced with a different affect, we can assess if it is a uniquely human aspect of language that infants initially overspecify their representations. We can also use a dog model to assess different theories about the kinds of language experience infants need to determine which aspects of speech define their word representations, and which do not. For example, one paper suggests that infants learn which phonetic distinctions are important to differentiate between words through experience with the distribution of phonemes in the words they have heard and learnt. The paper shows that hearing two contrasting sounds in different contexts improves infants' ability to utilize this contrast in a later word learning task (Thiessen, 2007). With a dog model, we could examine individual dogs' lexicons and see which phonetic contrasts their lexicon demonstrates in different contexts, and which contrasts are not present. We could then test dogs to see if they can detect the difference between two words that differ in the familiar contrast for the dogs, and whether they can detect a difference between two words containing an unfamiliar contrast, according to their lexicon structure. If learning phonemic contrasts relies on hearing sounds in many contexts, we would expect that dogs would succeed at the first task, and fail at the second.

One difficulty with these future studies is that they rely on researchers having an understanding of dogs' lexicon and their relative language exposure as compared to infants. There is currently no research that quantifies the extent of dogs' exposure to human speech. Additionally, no research has yet compared the structure and content of speech to dogs to that of speech to infants. This limits the current potential research questions to broadly and categorically researching the effect of a smaller lexicon and regular linguistic exposure, but not to examine the continuous effect of the structure and size of the lexicon or the amount of linguistic exposure. It is necessary to better understand dogs' lexicon and linguistic experience to test more specific questions. To better understand dogs' lexicons, some researchers are in the process of developing a linguistic survey, similar to the MacArthur-Bates Communicative Development Inventory (Fenson et al., 2007), that would allow dog owners to indicate which vocabulary items their dogs recognize, and add any extra words that are not on the list (Reeve & Jacques, 2019). These surveys would make it easier to do studies that compare high-vocabulary dogs' performance to lower-vocabulary dogs' performance in speech perception tasks to more precisely assess the effect of vocabulary size in the absence of a human linguistic system.

To better understand a dogs' language exposure as compared to infants, researchers could have several pet dogs wear a small recording device, such as a LENA (see Gilkerson & Richards, 2009, for more information about LENA use), and record spontaneous speech to these dogs over the course of several days. Researchers would then know both the amount of speech directed to dogs as compared to the amount of speech directed to infants in the course of a day, and information about the type of language used with dogs versus infants. The LENA could also help us get an understanding of the word forms dogs recognize. While owners can easily report commands that they say to their dog when determining what words they know, they may not remember all the words they often use when talking to their dog. It is possible that owners are repeating similar words very often to their dog that their dogs may recognize. The structure of dogs' word form vocabularies could contribute to the specificity of their representations.

In conclusion, pet dogs' language exposure and learned vocabulary is an asset for comparative speech perception research; however, further research is necessary to better understand the amount of language directed to dogs and how this compares with language directed to infants, as well as the amount and type of words dogs know and how that compares with infants' early vocabulary.

## *Implications for methodology and experimental design*

The last goal for this thesis was to better understand how to design methods for a comparative speech perception study for dogs and infants. An ideal comparative study allows for direct comparison between infant and dog results in order to accurately assess similarities and differences in performance. Below, I discuss the design choices made in the dissertation and the implications for future studies.

### Participant selection for dogs.

Prior to conducting the studies in Chapter 2, there were no clear benchmarks on the selection of canine participants for a speech perception study. The studies in this dissertation have led to an improved understanding of the number of participants necessary per study, as well as the effects of age, level of training, and psychiatric history.

***Number of participants.***

It is necessary to determine the ideal number of participants per study to maximize power and reduce the chance for false positive and false negative results. At first, the dog experiments used the same number of dog participants as we would use for an infant study, which is approximately 20-25 (see Newman, 2005; 2009). A power analysis conducted using the Hotelling-Lawley Trace on the overall data from Chapter 2 suggested that in a name-recognition-in-noise study, to see a main effect of Name versus Foil at a power of .8 and a Type 1 error rate at .05, it would be sufficient to test about 18 dogs. This validated the original choice to test 20 dogs in our study; however, for a more accurate prediction of the number of participants necessary to achieve a power of .8, more data were necessary. Once data collection for Chapter 3 was complete, I ran a second power analysis on the data from Chapters 2 and 3. This analysis, conducted using the Hotelling-Lawley Trace, suggests that in a name-recognition-in-noise study, to see a main effect of Name versus Foil at a power of .8 and a Type 1 error rate at .05, it would be necessary to test about 28 dogs. As such, the initial power estimate is a bit low, likely because dogs showed a fairly large effect in the nine-talker background noise studies from Chapter 2; this estimate is more conservative, as the dogs showed a smaller effect in single-talker background noise in Chapter 3.

Determining the number of dogs necessary to see an interaction between Name and Background Noise is harder, as the interaction varies depending on the signal-to-noise ratio and number of background talkers. In order to more accurately determine how many dogs are necessary to see an interaction, more data would be necessary in each of the types of background noise, and at each SNR level; currently, we only have about 20 participants in each of these conditions.

***Age.***

To ensure that dogs would have experience hearing their own name, the dogs in this dissertation were all at least 1 year old, and had heard their name for at least 10 months. While it was important to ensure that the older dogs did not have any significant hearing loss, as they may not have responded to our auditory stimuli, it was difficult for us to determine an upper age restriction, as different dogs have slightly different lifespans and begin to lose hearing at different ages. As such, owners were asked whether they noticed any hearing loss in their dog. This method was not perfect; we still tested and then dropped several dogs that appeared not to react to auditory stimuli (however, it is possible that these dogs could perceive the stimuli, but did not attend to the stimuli). Table 7, below, shows the oldest dogs in each study from this dissertation. While many older dogs successfully participated in our study, some failed to demonstrate any looking behavior and were dropped. There is no clear age demarcation at which dogs' performance began to decline; while some 15-year-old dogs successfully participated in the study, some 12-year-old dogs did not. As such, it would be beneficial to incorporate a hearing exam in the future for any dog studies of speech perception to better identify dogs that might have hearing loss and exclude them from studies.

Table 7: Oldest dog in each study

| Chapter | Study | Oldest dog (Only kept data) | Oldest dog (Including dropped data) |
| --- | --- | --- | --- |
| **Chapter 2** | Nine Talker 5 dB SNR | 12 years | 12 years |
| | Nine Talker 0 dB SNR | 11 years | 11 years |
| | Nine Talker -5 dB SNR | 11.25 years | 11.25 years |
| **Chapter 3** | Single Talker 0 dB SNR | 7 years | 12.5 years |
| | Single Talker 5 dB SNR (included in working dog analysis) | 11 years | 13 years |
| **Chapter 5** | Vowel/Consonant | 14 years | 14 years |
| | Mispronounce Only | 15 years | 15 years |

***Level of training.***

Other dog studies in the realm of speech perception and linguistic processing have used exceptional, highly-trained dogs with large learned vocabularies (Griebel & Oller, 2012; Kaminski et al., 2004; Pilley & Reid, 2011). The studies in this dissertation have shown that adult pet dogs as well as working dogs can be used for speech perception tasks, but the number of working dogs per group should be balanced; analyses from Chapter 2 and 3 indicate that working dogs show a stronger preference for their own name over a foil name than pet dogs. In Chapter 2, we did not control for the number of highly-trained dogs in each one of our studies. As such, the dogs in a more difficult signal-to-noise ratio, 0 dB SNR, appear to be doing better than the dogs in an easier signal-to-noise ratio, 5 dB SNR; however, this is potentially due to the large number of working dogs in the 0 dB SNR condition. Subsequent chapters controlled for the number of working dogs. Further studies should examine why working dogs show this stronger preference for their own name, and whether this improved performance in comparison to pet dogs extends to studies involving non-name-related auditory stimuli (for example, preference for most-often-heard language over an unfamiliar language). Determining whether working dogs' stronger preference is due to better attention or other factors like improved self-control could help with future studies that aim to disentangle attention effects in different speech perception tasks.

***Medication usage.***

Early in the process of testing dogs we anecdotally noticed that several dogs who were taking psychiatric medicine tended to be less attentive. As such, these dogs were not included in any of the studies in this dissertation. However, this observation is merely anecdotal and there is no current evidence indicating that dogs on psychiatric medicine are less attentive. Existing evidence of the effect of psychiatric medicine on human attention cannot shed light on potential effects in dogs, as some studies indicate that this medicine affects attention and cognitive processes (e.g. Rose, Simonotto, Spencer, & Ebmeier, 2006) and some indicate that it does not (e.g. Vermeeren et al., 1995). As such, further studies on the effect of psychiatric medicine on dogs' attention is necessary. One way to do so would be to compare dogs on psychiatric medicine to unmedicated dogs in the same study; for example, dogs on psychiatric medicine could be recruited and run in the nine-talker background noise at 5 dB SNR study from Chapter 2, and their performance could be compared to the existing dogs from the that study. If they do not show as strong of a name preference, or if they overall do not look at trials as long as the unmedicated dogs, it may indicate that the psychiatric medicine affects dogs' attention.

**Dog and infant stimuli.**

Overall, dogs and infants can generally be presented with the same stimuli. However, there are some notable differences in this dissertation in the manner in which the stimuli were recorded, and the type and structure of the stimuli used. Below, I discuss the changes made in recording and stimuli, their effectiveness, and the potential changes that could be made to make the dog and infant stimuli more similar for better comparison.

***Voice.***

In all the studies in this dissertation, dogs' names were presented in dog-directed speech rather than infant-directed speech. This was done because dogs have been shown to prefer dog-directed speech to adult-directed speech; however, there is no research about dogs' preferences for dog-directed speech in comparison to infant-directed speech. Dog-directed speech shares several qualities with infant-directed speech, including a similar higher-pitched register and a slower tempo in comparison to adult-directed speech (Burnham et al., 2002). However, infant-directed speech does have increased vowel hyperarticulation in comparison with pet-directed speech (Xu, Burnham, Kitamura, & Vollmer-Conna, 2013). This is especially important to note for experiments that involve vowel perception or word learning, like Chapter 5. While the target speech was not specifically tested for hyperarticulation in Chapter 5, the target speech was produced in a dog-directed manner. This could mean that there is less vowel hyperarticulation than there would be in infant-directed speech. However, the dogs in Chapter 5 were still able to distinguish between vowels. For more direct comparison, it would have been better to use infant-directed speech for both infants and dogs. Further research is necessary on dogs' preferences for dog- and infant-directed speech to address whether future experiments should utilize infant-directed speech for both dogs and infants, or whether dogs need to hear dog-directed speech for the best comparison.

***Stimuli type and structure.***

In general, the stimuli design can remain the same between dogs and infants, meaning that they will listen to the same number and same categories of auditory stimuli. However, since basic findings in dog research about their preferences for different types of auditory stimuli have not been established, some studies in this dissertation presented dogs with slightly different stimuli to examine the same questions as the original infant studies.

For example, infant studies examining name-in-noise perception have typically used a structure in which the infant is presented with their own name and three additional foil names, all in noise (Newman, 2005, 2009). Chapter 2 used a different structure to test dogs in which dogs heard their own name and a single foil, presented in both quiet and noise. This was done to ensure that dogs preferred their own name to a foil in quiet before assessing their performance in noise. Since this structure was so effective in Chapter 2, it was used in Chapter 3 as well.

While we have not presented dogs with their name and three foils in the presence of background noise, the second study in Chapter 5 uses the name-and-three-foils structure to test whether dogs consider a consonant-mispronounced version of their name to be similar to their own name. Dogs listen longer to the mispronounced version of their name in comparison to the foil names, which indicates that they are capable of showing a name preference given this structure.

**Testing procedure for dogs and infants.**

The studies in this dissertation all used the Headturn Preference Procedure (HPP). HPP is a commonly used infant research method to examine infants' preferences for different sounds and speech. We modified this method to test canines, which allows us to examine their understanding of different aspects of human speech and learn more about the perception of linguistic features that is shared with other species. Given dogs' social behavior and attention to humans around them, only a few changes were necessary from the original HPP design; we made edits to the booth lighting, the position of the dog and owner in the booth, and the attention getters before the trials. Coders were also made aware of how different dogs may behave in response to stimuli; for example, size sometimes affected dogs' looking behaviors (large dogs on the floor most often turned their whole body towards the audio source, while small dogs on the lap most often turned their head, like infants). Given the relatively few changes, it is still possible to make direct comparisons between dog and infant data using the same or similar stimuli.

It is not clear how well dog and infant performance would compare in other infant experimental paradigms. There is some evidence from comparative social cognition that dogs can be tested in a similar version of the Preferential Looking Procedure (Albuquerque et al., 2016; Racca et al., 2010) and the Expectancy

Violation paradigm (Adachi, Kuwahata, & Fujita, 2007; Kundey et al., 2010). However, studies from our lab that have tried to test dogs in Preferential Looking Procedure have not shown any results suggesting that dogs match auditory stimuli to visual stimuli in this paradigm; this matching did not occur even when our lab attempted a replication of an existing dog study using Preferential Looking Procedure (Yong & Ruffman, 2015). It is possible that our failure to show results is due to small changes in the apparatus of our preferential looking setup; for example, the screen we use to display visual stimuli is larger than the screens other researchers have used, and dogs may find our large, bright television screen aversive. As such, further studies are needed to identify the best apparatus for these testing procedures.

**Data analysis of dog and infant results.**

Many studies of infant speech perception have used mixed and within-subjects ANOVA to examine infants' preferences for different types of stimuli, including the infant studies to which Chapters 2, 3, and 5 were compared (Bouchon et al., 2015; Newman, 2005, 2009). Even other comparative developmental animal model studies have chosen to use the ANOVA due to its prevalence in this field (e.g., Bouchon & Toro, 2019). As such, the dog studies in this dissertation have mostly utilized mixed and within-subjects ANOVAs as well, for ideal comparison with the infant studies. However, the fields of psychology and linguistics are moving towards the use of mixed-effects models to analyze this type of research (Meteyard & Davies, 2020). Chapter 4 provides an example of how a mixed-effects model can be used to examine within-subjects infant data. Analyzing dog data could be done in the same manner.

It is important to note when selecting an analysis to use for comparative dog-infant data that dogs tend to have less variance overall in their performance than infants. As such it is always necessary to check for homoscedasticity, or equal variances, in the data before using an ANOVA to analyze the data. If the data is found to have unequal variances, it is possible to do a log-transformation on the outcome variable and re-check for homoscedasticity. If the variances are still unequal, the best option would be to use a mixed-effects model.

### *Future directions to improve the dog model*

This dissertation presented example studies and results utilizing a dog model for developmental speech perception to suggest that dog models could be used more widely to answer questions that would be otherwise unanswerable by testing only infants and adults. Future studies are necessary to better understand dogs' underlying systems of speech perception, their language exposure, and to find other useful testing methodologies for dogs, such that researchers can determine the kinds of research questions that a dog model would be best suited to answer.

One way to improve the dog model is to do more fundamental research in order to help better understand and identify more uses for the dog model. As discussed above, while current research in the underlying auditory, attention, and linguistic systems of dogs and infants allows for broad questions about the primary systems involved in specific speech perception tasks, there is still much that is not known about these systems in dogs. A clearer understanding of how these systems function in dogs and how they compare with the analogous infant systems would allow researchers to use the dog model for more complex comparative questions.

Another way to improve the dog model is to characterize the nature of dogs' linguistic exposure and to determine their lexicon structure. This will allow researchers to ask more precise questions about the amount of linguistic exposure necessary for certain speech perception abilities to develop. One study possibility would be to use a recording device to determine how much speech dogs hear on a regular basis and how this correlates with dogs' knowledge of words and commands.

Lastly, researchers could also explore different testing methodologies that can be used with both dogs and infants. All the studies in this dissertation utilize the Headturn Preference Procedure (HPP). This method is very versatile, but it is not the only method used to test questions in infant speech perception. One of the main limitations of classic HPP is the inability to display images and videos to the participant in addition to the auditory stimuli. Further studies are necessary to assess whether dogs can be reliably tested in other methods, like the preferential looking procedure, central fixation procedure, and expectancy-violation procedure, and whether their results are comparable to those of infants. Understanding the other kinds of methods that can be used with both dogs and infants will allow researchers to test many different types of questions in speech perception, including questions that involve visual stimuli (e.g., questions about word-object matching).

### *General conclusion*

In conclusion, the studies in this dissertation provide evidence that dogs can be a useful comparative model for developmental speech perception. Particularly, dogs are well-suited for broad questions about the contributions of the system in speech perception tasks, as well as the role of language experience in the development of speech perception abilities. To ask more specific questions, further research is necessary to better characterize dogs' language experience as well as the function of dogs' auditory, attention, and linguistic systems. Studies are also needed to better understand the comparative paradigms in which researchers can test dogs, and to identify any changes that must be made from the original human paradigm in order for dogs to successfully participate. Better fundamental understanding of the processes underlying speech perception in dogs along with improved methodological understanding of comparative infant-dog testing paradigms can allow researchers to better utilize a comparative dog model for a wider variety of developmental speech perception questions.

# Bibliography

Adachi, I., Kuwahata, H., & Fujita, K. (2007). Dogs recall their owner's face upon hearing the owner's voice. *Animal Cognition*, *10*(1), 17–21. https://doi.org/10.1007/s10071-006-0025-8

Adams, C. L., Molfese, D. L., & Betz, J. C. (1987). Electrophysiological correlates of categorical speech perception for voicing contrasts in dogs. *Developmental Neuropsychology, 3*(3–4), 175–189. https://doi.org/10.1080/87565648709540375

Albuquerque, N., Guo, K., Wilkinson, A., Resende, B., & Mills, D. S. (2018). Mouth-licking by dogs as a response to emotional stimuli. *Behavioural Processes, 146,* 42–45. https://doi.org/10.1016/j.beproc.2017.11.006

Albuquerque, N., Guo, K., Wilkinson, A., Savalli, C., Otta, E., & Mills, D. (2016). Dogs recognize dog and human emotions. *Biology Letters*, *12*(1), 20150883. https://doi.org/10.1098/rsbl.2015.0883

Alterisio, A., Scandurra, A., Eatherington, C. J., Marinelli, L., D'Aniello, B., & Mongillo, P. (2019). You can't see, when I do: A study on social attention in guide dogs. *Applied Animal Behaviour Science, 218,* 104824. https://doi.org/10.1016/j.applanim.2019.06.005

Andics, A., Gábor, A., Gácsi, M., Faragó, T., Szabó, D., & Miklósi, Á. (2016). Neural mechanisms for lexical processing in dogs. *Science*, *353*(6303), 1030–1032. https://doi.org/10.1126/science.aaf3777

Andics, A., Gácsi, M., Faragó, T., Kis, A., & Miklósi, Á. (2014). Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. *Current Biology*, *24*(5), 574–578. https://doi.org/10.1016/j.cub.2014.01.058

Andics, A., & Miklósi, Á. (2018). Neural processes of vocal social perception: Dog-human comparative fMRI studies. *Neuroscience and Biobehavioral Reviews*, *85*(November 2017), 54–64. https://doi.org/10.1016/j.neubiorev.2017.11.017

Anrep, G. V. (1920). Pitch discrimination in the dog. *Journal of Experimental Medicine*, *53*(6), 367–385.

Appleyard, D., & Lintell, M. (1972). The environmental quality of city streets: The residents' viewpoint. *Journal of the American Planning Association*, *38*(2), 84–101. https://doi.org/10.1080/01944367208977410

Ashmead, D. H., Clifton, R. K., & Perris, E. E. (1987). Precision of auditory localization in human infants. *Developmental Psychology*, *23*(5), 641–647. https://doi.org/10.1037/0012-1649.23.5.641

Ashmead, D. H., Clifton, R. K., & Reese, E. P. (1986). Development of auditory localization in dogs: Single source and precedence effect sounds. *Developmental Psychobiology*, *19*(2), 91–103. https://doi.org/10.1002/dev.420190202

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. *Child Development*, *52*(4), 1135. https://doi.org/10.1111/j.1467-8624.1981.tb03159.x

Bach, J. P., Lüpke, M., Dziallas, P., Wefstaedt, P., Uppenkamp, S., Seifert, H., &

Nolte, I. (2016). Auditory functional magnetic resonance imaging in dogs - normalization and group analysis and the processing of pitch in the canine auditory pathways. *BMC Veterinary Research*, *12*(1), 1–9. https://doi.org/10.1186/s12917-016-0660-5

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, *32*(1), 159–173. https://doi.org/10.1017/S0305000904006567

Bargones, J. Y., & Werner, L. A. (1994). Adults listen selectively; infants do not. *Psychological Science*, *5*(3), 170–174. https://doi.org/10.1111/j.1467-9280.1994.tb00655.x

Bargones, J. Y., Werner, L. A., & Marean, G. C. (1995). Infant psychometric functions for detection: Mechanisms of immature sensitivity. *Journal of the Acoustical Society of America*, *98*(1), 99–111. https://doi.org/10.1121/1.414446

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2), 45–53. https://doi.org/10.1016/j.cognition.2004.06.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Ben-Aderet, T., Gallego-Abenza, M., Reby, D., & Mathevon, N. (2017). Dog-directed speech: why do we use it and do dogs pay attention to it? *Proceedings of the Royal Society B: Biological Sciences*, *284*(1846), 20162429. https://doi.org/10.1098/rspb.2016.2429

Bendixen, A., Háden, G. P., Németh, R., Farkas, D., Török, M., & Winkler, I. (2015). Newborn infants detect cues of concurrent sound segregation. *Developmental Neuroscience*, *37*(2), 172–181. https://doi.org/10.1159/000370237

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International*, *5*(9–10), 341–347.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations. *Psychological Science*, *16*(6), 451–459. https://doi.org/10.1111/j.0956-7976.2005.01556.x

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298–304. https://doi.org/10.1111/j.0956-7976.2005.01531.x

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, *18*(4), 587–598. https://doi.org/10.1111/desc.12242

Bouchon, C., & Toro, J. M. (2019). Is the consonant bias specifically human? Long–Evans rats encode vowels better than consonants in words. *Animal Cognition*, *22*(5), 839–850. https://doi.org/10.1007/s10071-019-01280-3

Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, *67*(2), 297. https://doi.org/10.2307/415108

Brandt, M., & Bitzer, J. (2014). Automatic detection of hum in audio signals. *JEAS*, *62*(9), 584–595.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44. https://doi.org/10.1016/S0010-0277(01)00122-6

Bugannim, Y., Roth, D. A. E., Zechoval, D., & Kishon-Rabin, L. (2019). Training of speech perception in noise in pre-lingual hearing impaired adults with cochlear implants compared with normal hearing adults. *Otology and Neurotology*, *40*(3), E316–E325. https://doi.org/10.1097/MAO.0000000000002128

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, *296*(5572), 1435. https://doi.org/10.1126/science.1069587

Cavalli, C. M., Carballo, F., Dzik, M. V., Underwood, S., & Bentosela, M. (2018). Are animal-assisted activity dogs different from pet dogs? A comparison of their sociocognitive abilities. *Journal of Veterinary Behavior: Clinical Applications and Research*, *23*, 76–81. https://doi.org/10.1016/j.jveb.2017.12.001

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. https://doi.org/10.1038/1561

Clark, F. E. (2011). Great ape cognition and captive care: Can cognitive challenges enhance well-being? *Applied Animal Behaviour Science*, *135*, 1–12. https://doi.org/10.1016/j.applanim.2011.10.010

Clarkson, M. G., Clifton, R. K., Swain, I. U., & Perris, E. E. (1989). Stimulus duration and repetition rate influence newborns' head orientation toward sound. *Developmental Psychobiology*, *22*(7), 683–705. https://doi.org/10.1002/dev.420220704

Clifton, R. K., Morrongiello, B. A., Kulig, J. W., & Dowd, J. M. (1981). Newborns' orientation toward sound: Possible implications for cortical development. *Child Development*, *52*(3), 833–838.

Coch, D., Sanders, L. D., & Neville, H. J. (2005). An event-related potential study of selective auditory attention in children and adults. *Journal of Cognitive Neuroscience*, *17*(4), 605–622. https://doi.org/10.1162/0898929053467631

Cohen, J. A., & Fox, M. W. (1976). Vocalizations in wild canids and possible effects of domestication. *Behavioural Processes*, *1*(1), 77–92. https://doi.org/10.1016/0376-6357(76)90008-5

Coplan, J. (1993). Early language milestone scale-2. Austin, TX: Pro-Ed.

Coren, S. (2009). How dogs think. In *The American Psychological Association's 117th Annual Convention*. Toronto, ON.

Cuaya, L. V, Hernandez-Perez, R., & Concha, L. (2016). Our faces in the dog's brain: Functional imaging reveals temporal cortex activation during perception of human faces. *PloS ONE*, *11*(3), e0149431.

https://doi.org/10.1371/journal.pone.0149431

Culling, J. F., & Stone, M. A. (2017). Energetic masking and masking release. In J. C. Middlebrooks, J. Z. Simon, A. F. Popper, & R. R. Fay (Eds.), *The auditory system at the cocktail party* (pp. 41–73). Springer, Cham.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics, 21*(1–2), 103–108. https://doi.org/10.1016/s0095-4470(19)31323-3

Cutler, A., Sebastian-Galles, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory and Cognition, 28*(5), 746–755. https://doi.org/10.3758/BF03198409

Darwin, C. J. (2008). Listening to speech in the presence of other sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1011–1021. https://doi.org/10.1098/rstb.2007.2156

Decasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *American Association for the Advancement of Science, 208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

Delle Luche, C., Poltrock, S., Goslin, J., New, B., Floccia, C., & Nazzi, T. (2014). Differential processing of consonants and vowels in the auditory modality: A cross-linguistic study. *Journal of Memory and Language*, *72*, 1–15. https://doi.org/10.1016/j.jml.2013.12.001

DePaolis, R. A., Keren-Portnoy, T., & Vihman, M. (2016). Making sense of infant familiarity and novelty responses to words at lexical onset. *Frontiers in*

*Psychology*, *7*, 715. https://doi.org/10.3389/fpsyg.2016.00715

Dorey, N. R., Udell, M. A. R., & Wynne, C. D. L. (2009). Breed differences in dogs sensitivity to human points: A meta-analysis. *Behavioural Processes*, *81*(3), 409–415. https://doi.org/10.1016/j.beproc.2009.03.011

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD rating scale—IV: Checklists, norms, and clinical interpretation*. Guilford Press.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *The Journal of the Acoustical Society of America*, *114*(1), 368–379. https://doi.org/10.1121/1.1577562

Eimas, P. D. (1996). The perception and representation of speech by infants. In James L. Morgan & K. Demuth (Eds.), *Signal to Syntax* (1st ed.). New York, NY.

Erickson, L. C., & Newman, R. S. (2017). Influences of background noise on infants and children. *Current Directions in Psychological Science*, *26*(5), 451–457. https://doi.org/10.1177/0963721417709087

Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*(3), 254–260. https://doi.org/10.1111/j.1467-9280.2007.01885.x

Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist*, *59*(2), 77–92. https://doi.org/10.1037/0003-066X.59.2.77

Evans, G. W., & Kantrowitz, E. (2002). Socioeconomic status and health: The potential role of environmental risk exposure. *Annual Review of Public Health*, *23*(1), 303–331. https://doi.org/10.1146/annurev.publhealth.23.112001.112349

Faragó, T., Pongrácz, P., Range, F., Virányi, Z., & Miklósi, Á. (2010). “The bone is mine”: affective and referential aspects of dog growls. *Animal Behaviour*, *79*(4), 917–925. https://doi.org/10.1016/j.anbehav.2010.01.005

Faragó, T., Townsend, S., & Range, F. (2014). The information content of wolf (and dog) social communication. In *Biocommunication of Animals* (pp. 41–52). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-7414-8

Feddersen-Petersen, D. U. (2000). Vocalization of European wolves (Canis lupus lupus L.) and various dog breeds (Canis lupus f. fam.). *Archives Animal Breeding*, *43*(4), 387–397. https://doi.org/10.5194/aab-43-387-2000

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, S. J., & Bates, E. (2007). *MacArthur-Bates communicative development inventories*. Paul H. Brookes Publishing Co.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, *88*(4), 1725–1736. https://doi.org/10.1121/1.400247

Finitzo-Hieber, T., & Tillman, T. W. (1978). Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children. *Journal of Speech and Hearing Research*, *21*(3), 440–458. https://doi.org/10.1044/jshr.2103.440

Floccia, C., Nazzi, T., Delle Luche, C., Poltrock, S., & Goslin, J. (2014). English-learning one- to two-year-olds do not show a consonant bias in word learning. *Journal of Child Language*, *41*(5), 1085–1114. https://doi.org/10.1017/S0305000913000287

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*. https://doi.org/10.1017/s0305000916000209

Fugazza, C., & Miklósi, Á. (2014). Deferred imitation and declarative memory in domestic dogs. *Animal Cognition*, *17*(2), 237–247. https://doi.org/10.1007/s10071-013-0656-5

Fugazza, C., & Miklósi, Á. (2020). Depths and limits of spontaneous categorization in a family dog. *Scientific Reports*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-59965-6

Gácsi, M., Miklód, Á., Varga, O., Topál, J., & Csányi, V. (2004). Are readers of our face readers of our minds? Dogs (Canis familiaris) show situation-dependent recognition of human's attention. *Animal Cognition*, *7*(3), 144–153. https://doi.org/10.1007/s10071-003-0205-8

Gale, A., & Lynn, R. (1972). A developmental study of attention. *The British Journal* *of Educational Psychology*, *42*(3), 260–266. https://doi.org/10.1111/j.2044-8279.1972.tb00719.x

Galinsky, E. (2011). *Mind in the making: Experiments in children's learning*. New Screen Concepts (Firm).

Gervain, J., & Werker, J. F. (2008). How infant speech perception contributes to language acquisition. *Linguistics and Language Compass*, *2*(6), 1149–1170. https://doi.org/10.1111/j.1749-818X.2008.00089.x

Gilkerson, J., & Richards, J. A. (2009). *The LENA natural language study*. *LENA Foundation Technical Report*.

Gomes, H. (2000). The development of auditory attention in children. *Frontiers in Bioscience*, *5*(1), 108–120. https://doi.org/10.2741/Gomes

Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., & Cowan, N. (2000). The development of auditory attention in children. *Frontiers in Bioscience*, *5*(1), 108–120. https://doi.org/10.2741/Gomes

Griebel, U., & Oller, D. K. (2012). Vocabulary learning in a Yorkshire terrier: Slow mapping of spoken words. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0030182

Haines, M. M., Stansfeld, S. A., Head, J., & Job, R. F. S. (2002). Multilevel modelling of aircraft noise on performance tests in schools around Heathrow Airport London. *Journal of Epidemiology and Community Health*, *56*(2), 139–144. https://doi.org/10.1136/jech.56.2.139

Hallé, P. A., & Boysson-Bardies, B. de. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development*, *17*(2), 119–129. https://doi.org/10.1016/0163-6383(94)90047-7

Hare, B., Brown, M., Williamson, C., & Tomasello, M. (2002). The domestication of social cognition in dogs. *Science*, *298*(5598), 1634–1636. https://doi.org/10.1126/science.1072702

Hare, B., Call, J., & Tomasello, M. (2010). Communication of food location between human and dog (Canis familiaris). *Evolution of Communication*, *2*(1), 137–159. https://doi.org/10.1075/eoc.2.1.06har

Hart, B., & Risley, R. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Paul H. Brooks.

Hayashi, A., Tamekawa, Y., & Kiritani, S. (2001). Developmental change in auditory preferences for speech stimuli in Japanese infants. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1189–1200. https://doi.org/10.1044/1092-4388(2001/092)

Heffner, H. E. (1978). Effect of auditory cortex ablation on localization and discrimination of brief sounds. *Journal of Neurophysiology*, *41*(4), 963–976. https://doi.org/10.1152/jn.1978.41.4.963

Heffner, H. E. (1983). Hearing in large and small dogs: Absolute thresholds and the size of the tympanic membrane. *Behavioral Neuroscience*, *97*(2), 310–318.

Heffner, R. S., & Heffner, H. E. (1992). Evolution of sound localization in mammals. In D. B. Webster, A. N. Popper, & R. R. Fay (Eds.), *The evolutionary biology of hearing* (pp. 691–715). New York, NY: Springer.

Herman, L. M. (1986). Cognition and language competencies of bottlenosed dolphins. In R. J. Schusterman, J. A. Thomas, & F. G. Wood (Eds.), *Dolphin cognition and behavior: A comparative approach* (pp. 221–253).

Hochmann, J. R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science*, *14*(6), 1445–1458. https://doi.org/10.1111/j.1467-7687.2011.01089.x

Højen, A., & Nazzi, T. (2016). Vowel bias in Danish word-learning: Processing biases are language-specific. *Developmental Science*, *19*(1), 41–49. https://doi.org/10.1111/desc.12286

Hollich, G. (2006). Combining techniques to reveal emergent effects in infants' segmentation, word learning, and grammar. *Language and Speech*, *49*(Pt 1), 3–19. https://doi.org/10.1177/00238309060490010201

Holmes, D. J., & Austad, S. N. (1995). Birds as animal models for the comparative biology of aging: A prospectus. *Journals of Gerontology - Series A: Biological Sciences and Medical Sciences*, *50*(2), B59–B66. https://doi.org/10.1093/gerona/50A.2.B59

Horowitz, A. (2009). Attention to attention in domestic dog (Canis familiaris) dyadic play. *Animal Cognition*, *12*(1), 107–118. https://doi.org/10.1007/s10071-008-0175-y

Horowitz, A., & Bekoff, M. (2007). Naturalizing anthropomorphism: Behavioral prompts to our humanizing of animals. *Anthrozoös*, *20*(1), 23–35. https://doi.org/10.2752/089279307780216650

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. https://doi.org/10.1080/15250000701795598

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1570–1582. https://doi.org/10.1037/0096-1523.26.5.1570

Huber, L., Racca, A., Scaf, B., Virányi, Z., & Range, F. (2013). Discrimination of familiar human faces in dogs (Canis familiaris). *Learning and Motivation*, *44*(4), 258–269. https://doi.org/10.1016/j.lmot.2013.04.005

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (pp. 69–95). Westport, CT: Ablex.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, *45*(3), 337–374. https://doi.org/10.1016/S0010-0285(02)00500-5

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science, 14*(5), 1002–1011. https://doi.org/10.1111/j.1467-7687.2011.01052.x

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*(3), 402–420.

https://doi.org/10.1006/jmla.1993.1022

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*, 630–645. Retrieved from http://repositorio.unan.edu.ni/2986/1/5624.pdf

Kagan, J., & Lewis, M. (1965). Studies of attention in the human infant. *Merrill-Palmer Quarterly of Behavior and Development*, *11*(2), 95–127.

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: Evidence for "fast mapping." *Science*, *304*(5677), 1682–1683. https://doi.org/10.1126/science.1097859

Keidel, J. L., Jenison, R. L., Kluender, K. R., & Seidenberg, M. S. (2007). Does grammar constrain statistical learning? *Psychological Science*, *18*(10), 922. https://doi.org/10.1111/j.1467-9280.2007.02001.x

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference for testing auditory perception. *Infant Behavior and Development*, *18*, 111–116.

Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *Journal of the Acoustical Society of America*, *95*(6), 3475–3480. https://doi.org/10.1121/1.410023

Kluender, K. R., Lotto, A. J., & Holt, L. L. (2006). Contributions of nonhuman animal models to understanding human speech perception. In S. Greenberg & W. Ainsworth (Eds.), *Listening to speech: An auditory perspective* (Vol. 33, pp. 369–374). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science*, *1*(1), 70–73. https://doi.org/10.1111/j.1467-9280.1990.tb00070.x

Kuchinsky, S. E., Ahlstrom, J. B., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2014). Speech-perception training for older adults with hearing loss impacts word recognition and effort. *Psychophysiology*, *51*(10), 1046–1057. https://doi.org/10.1111/psyp.12242

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America*, *63*(3), 905–917. https://doi.org/10.1121/1.381770

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608. https://doi.org/10.1109/ECTC.2006.1645911

Kundey, S. M. A., de Los Reyes, A., Taglang, C., Baruch, A., & German, R. (2010). Domesticated dogs' (Canis familiaris) use of the solidity principle. *Animal Cognition*, *13*(3), 497–505. https://doi.org/10.1007/s10071-009-0300-6

Kutsumi, A., Nagasawa, M., Ohta, M., & Ohtani, N. (2012). Importance of puppy training for future behavior of the dog. *Journal of Veterinary Medical Science*, *75*(2), 141–149. https://doi.org/10.1292/jvms.12-0008

Ladefoged, P. (2001). *Vowel and consonants: An introduction to the sounds of language*. Oxford: Blackwell.

Lapierre, M. A., Piotrowski, J. T., & Linebarger, D. L. (2012). Background television in the homes of US children. *Pediatrics*, *130*(5), 839–846. https://doi.org/10.1542/peds.2011-2581

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*. https://doi.org/10.1037/h0020279

Lipman, E. A., & Grassi, J. R. (1942). Comparative auditory sensitivity of man and dog. *The American Journal of Psychology*, *55*(1), 84–89.

Lit, L., Schweitzer, J. B., Iosif, A. M., & Oberbauer, A. M. (2010). Owner reports of attention, activity, and impulsivity in dogs: A replication study. *Behavioral and Brain Functions*, *6*, 1–10. https://doi.org/10.1186/1744-9081-6-1

Lord, K. (2013). A Comparison of the sensory development of wolves (Canis lupus lupus) and dogs (Canis lupus familiaris). *Ethology*, *119*(2), 110–120. https://doi.org/10.1111/eth.12044

Lyn, H., & Savage-Rumbaugh, E. S. (2000). Observational word learning in two bonobos (Pan paniscus): Ostensive and non-ostensive contexts. *Language and Communication*, *20*(3), 255–273. https://doi.org/10.1016/s0271-5309(99)00026-9

Mallikarjun, A., Shroads, E., & Newman, R. S. (2019). The cocktail party effect in the domestic dog (Canis familiaris). *Animal Cognition*. https://doi.org/10.1007/s10071-019-01255-4

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*(5), 314–317.

https://doi.org/10.1111/j.1467-9280.1995.tb00517.x

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272. https://doi.org/10.1016/j.jml.2007.03.005

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Marshall-Pescini, S., Passalacqua, C., Barnard, S., Valsecchi, P., & Prato-Previde, E. (2009). Agility and search and rescue training differently affects pet dogs' behaviour in socio-cognitive tasks. *Behavioural Processes*, *81*(3), 416–422. https://doi.org/10.1016/j.beproc.2009.03.015

Marshall-Pescini, S., Valsecchi, P., Petak, I., Accorsi, P. A., & Previde, E. P. (2008). Does training make you smarter? The effects of training on dogs' performance (Canis familiaris) in a problem solving task. *Behavioural Processes*, *78*(3), 449–454. https://doi.org/10.1016/j.beproc.2008.02.022

Mather, E., & Plunkett, K. (2011). Mutual exclusivity and phonological novelty constrain word learning at 16 months. *Journal of Child Language*, *38*(5), 933–950. https://doi.org/10.1017/S0305000910000401

McAlexander, T. P., Gershon, R. R. M., & Neitzel, R. L. (2015). Street-level noise in an urban setting: Assessment and contribution to personal exposure. *Environmental Health: A Global Access Science Source*, *14*(1). https://doi.org/10.1186/s12940-015-0006-y

McComb, K., Shannon, G., Sayialel, K. N., & Moss, C. (2014). Elephants can determine ethnicity, gender, and age from acoustic cues in human voices. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(14), 5433–5438. https://doi.org/10.1073/pnas.1321543111

McKinley, J., & Sambrook, T. D. (2000). Use of human-given cues by domestic dogs (Canis familiaris) and horses (Equus caballus). *Animal Cognition*, *3*, 13–22. https://doi.org/10.1007/s100710050046

McMillan, B. T. M., & Saffran, J. R. (2016). Learning in complex environments: The effects of background speech on early word learning. *Child Development*, *87*(6), 1841–1855. https://doi.org/10.1111/cdev.12559

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838), 631. https://doi.org/10.1126/science.1144073

Mehler, J., Bertoncini, J., Barriere, M., & Jassik Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, *7*(5), 491–497. https://doi.org/10.1068/p070491

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101–116). Mahwah, NJ: Erlbaum.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Merola, I., Prato-Previde, E., & Marshall-Pescini, S. (2012). Dogs' social referencing towards owners and strangers. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0047653

Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, *54*(3/4), 1–129. https://doi.org/10.2307/1166130

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for LMMs: Best practice guidance for LMMs. *Journal of Memory and Language*, *112*, 104092. https://doi.org/DOI 10.17605/OSF.IO/BFQ39

Miklösi, Á., Polgárdi, R., Topál, J., & Csányi, V. (1998). Use of experimenter-given cues in dogs. *Animal Cognition*, *1*(2), 113–121. https://doi.org/10.1007/s100710050016

Mitchell, R. W. (2001). Americans' talk to dogs: Similarities and differences with talk to infants. *Research on Language and Social Interaction*, *34*(2), 183–210. https://doi.org/10.1207/S15327973RLSI34-2

Mongillo, P., Bono, G., Regolin, L., & Marinelli, L. (2010). Selective attention to humans in companion dogs , Canis familiaris. *Animal Behaviour*, *80*(6), 1057–1063. https://doi.org/10.1016/j.anbehav.2010.09.014

Morrongiello, B. A. (1988). Infants' localization of sounds along the horizontal axis: Estimates of minimum audible angle. *Developmental Psychology*, *24*(1), 8–13. https://doi.org/10.1037/0012-1649.24.1.8

Morrongiello, B. A., & Clifton, R. K. (1984). Effects of sound frequency on behavioral and cardiac orienting in newborn and five-month-old infants. *Journal of Experimental Child Psychology*, *38*(3), 429–446. https://doi.org/10.1016/0022-0965(84)90086-9

Morrongiello, B. A., Fenwick, K. D., Hillier, L., & Chance, G. (1994). Sound localization in newborn human infants. *Developmental Psychobiology*, *27*(8), 519–538. https://doi.org/10.1002/dev.420270805

Morrongiello, B. A., & Rocca, P. T. (1987). Infants' localization of sounds in the horizontal plane: effects of auditory and visual cues. *Child Development*, *58*(4), 918–927. https://doi.org/10.1111/j.1467-8624.1987.tb01429.x

Morse, P. A., Molfese, D., Laughlin, N. K., Linnville, S., & Wetzel, F. (1987). Categorical perception for voicing contrasts in normal and lead-treated rhesus monkeys: Electrophysiological indices. *Brain and Language*, *30*(1), 63–80. https://doi.org/10.1016/0093-934X(87)90028-9

Most popular U.S. pet names of 2019. (2019). Retrieved from https://www.rover.com/blog/dog-names/

Muir, D. W., & Clifton, R. K. (1985). Infants' orientation to the location of sound sources. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 171–194). Westport, CT: Ablex Publishing.

Muir, D. W., Clifton, R. K., & Clarkson, M. G. (1989). The development of a human auditory localization response: a U-shaped function. *Canadian Journal of Psychology*, *43*(2), 199. https://doi.org/10.1037/h0084220

Musiek, F. E., Shinn, J. B., Jirsa, R., Bamiou, D. E., Baran, J. A., & Zaida, E. (2005). GIN (Gaps-In-Noise) test performance in subjects with confirmed central auditory nervous system involvement. *Ear and Hearing*, *26*(6), 608–618. https://doi.org/10.1097/01.aud.0000188069.80699.41

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*(1), 13–30. https://doi.org/10.1016/j.cognition.2004.10.005

Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, *102*(4), 522–537. https://doi.org/10.1016/j.jecp.2008.05.003

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *Journal of the Acoustical Society of America*, *98*(4), 1909–1920. https://doi.org/10.1121/1.414458

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*(2), 203–230. https://doi.org/10.1418/10879

Newman, R. S. (2005). The cocktail party effect in infants revisited: listening to one's name in noise. *Developmental Psychology*, *41*(2), 352–362. https://doi.org/10.1037/0012-1649.41.2.352

Newman, R. S. (2009). Infants' listening in multitalker environments: Effect of the number of background talkers. *Attention, Perception, & Psychophysics*, *71*(4), 822–836. https://doi.org/10.3758/APP

Newman, R. S., & Jusczyk, P. W. (1996). The cocktail party effect in infants. *Perception & Psychophysics*, *58*(8), 1145–1156. https://doi.org/10.3758/BF03207548

Newman, R. S., Morini, G., Ahsan, F., & Kidd, G. (2015). Linguistically-based informational masking in preschool children. *The Journal of the Acoustical Society of America*, *138*(1), EL93–EL98. https://doi.org/10.1121/1.4921677

Newman, R. S., Rowe, M. L., & Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, *43*(5), 1158–1173. https://doi.org/10.1017/S0305000915000446

Newman, R. S., Shroads, E., Morini, G., Johnson, E. K., Onishi, K. H., & Tincoff, R. (2019). BITTSy: Behavioral infant & toddler testing system. Retrieved from http://go.umd.edu/BITTSy

Nike. (2016). Speech spectrum shaped noise. Matlab File Exchange.

Nishibayashi, L. L., & Nazzi, T. (2016). Vowels, then consonants: Early bias switch in recognizing segmented word forms. *Cognition*, *155*, 188–203. https://doi.org/10.1016/j.cognition.2016.07.003

Nozza, R. J. (2005). Infant speech-sound discrimination testing: Effects of stimulus intensity and procedural model on measures of performance. *The Journal of the Acoustical Society of America*, *81*(6), 1928–1939. https://doi.org/10.1121/1.394757

Nozza, R. J., Rossman, R. N., Bond, L. C., & Miller, S. L. (1990). Infant speech-sound discrimination in noise. *The Journal of the Acoustical Society of America*, *87*(1), 339–350. https://doi.org/10.1121/1.399301

Nozza, R. J., Rossman, R. N. F., & Bond, L. C. (1991). Infant-adult differences in unmasked thresholds for the discrimination of consonant-vowel syllable pairs. *International Journal of Audiology*, *30*(2), 102–112. https://doi.org/10.3109/00206099109072875

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.

Olsho, L. W., Koch, E. G., Carter, E. A., Halpin, C. F., & Spetner, N. B. (1988). Pure-tone sensitivity of human infants. *The Journal of the Acoustical Society of America*, *84*(4), 1316–1324. https://doi.org/10.1121/1.2023630

Olsho, L. W., Koch, E. G., & Halpin, C. F. (1987). Level and age effects in infant frequency discrimination. *The Journal of the Acoustical Society of America*, *82*(2), 454–464. https://doi.org/10.1121/1.395446

Olsho, L. W., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982). Auditory frequency discrimination in infancy. *Developmental Psychology*, *18*(5), 721–726. https://doi.org/10.1037/0012-1649.18.5.721

Patterson, F. G., & Cohn, R. H. (1990). Language acquisition by a lowland gorilla: Koko’s first ten years of vocabulary development. *Word*, *41*(2), 97–143. https://doi.org/10.1080/00437956.1990.11435816

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and*

*Development*, *15*(3), 325–345. https://doi.org/10.1016/0163-6383(92)80003-D

Perez, C. A., Engineer, C. T., Jakkamsetti, V., Carraway, R. S., Perry, M. S., & Kilgard, M. P. (2013). Different timescales for the neural coding of consonant and vowel sounds. *Cerebral Cortex*, *23*(3), 670–683. https://doi.org/10.1093/cercor/bhs045

Phillips, D. P., Comeau, M., & Andrus, J. N. (2011). Auditory temporal gap detection in children with and without auditory processing disorder. *Journal of the American Academy of Audiology*, *21*(6), 404–408. https://doi.org/10.3766/jaaa.21.6.5

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., & Mossop, J. E. (1997). Detection of silent intervals between noises activating different perceptual channels: Some properties of "central" auditory gap detection. *The Journal of the Acoustical Society of America*, *101*(6), 3694–3705. https://doi.org/10.1121/1.419376

Pilley, J. W. (2013). Border collie comprehends sentences containing a prepositional object, verb, and direct object. *Learning and Motivation*, *44*(4), 229–240. https://doi.org/10.1016/j.lmot.2013.02.003

Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*(2), 184–195. https://doi.org/10.1016/j.beproc.2010.11.007

Polka, L., Rvachew, S., & Molnar, M. (2008). Speech perception by 6- to 8-month-olds in the presence of distracting sounds. *Infancy*, *13*(5), 421–439.

https://doi.org/10.1080/1

Poltrock, S., & Nazzi, T. (2015). Consonant/vowel asymmetry in early word form recognition. *Journal of Experimental Child Psychology*, *131*, 135–148. https://doi.org/10.1016/j.jecp.2014.11.011

Proops, L., & Mccomb, K. (2012). Cross-modal individual recognition in domestic horses (equus caballus) extends to familiar humans. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1741), 3131–3138. https://doi.org/10.1098/rspb.2012.0626

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Racca, A., Amadei, E., Ligout, S., Guo, K., Meints, K., & Mills, D. (2010). Discrimination of human and dog faces and inversion responses in domestic dogs (Canis familiaris). *Animal Cognition*, *13*(3), 525–533. https://doi.org/10.1007/s10071-009-0303-3

Range, F., & Virányi, Z. (2013). Social learning from humans or conspecifics: Differences and similarities between wolves and dogs. *Frontiers in Psychology*, *4*, 868. https://doi.org/10.3389/fpsyg.2013.00868

Ratcliffe, V. F., & Reby, D. (2014). Orienting asymmetries in dogs' responses to different communicatory components of human speech. *Current Biology*, *24*(24), 2908–2912. https://doi.org/10.1016/j.cub.2014.10.030

Reed, P., Howell, P., Sackin, S., Pizzimenti, L., & Rosen, S. (2003). Speech perception in rats: Use of duration and rise time cues in labeling of affricate/fricative sounds. *Journal of the Experimental Analysis of Behavior*, *80*(2), 205–215. https://doi.org/10.1901/jeab.2003.80-205

Reeve, C., & Jacques, S. (2019). Knowledge of spoken words by domestic dogs: A new instrument for use with dog owners. Presented at the *Society for Research in Child Development*. Baltimore, MD.

Rhodes, G., Geddes, K., Jeffery, L., Dziurawiec, S., & Clark, A. (2002). Are average and symmetric faces attractive to infants? Discrimination and looking preferences. *Perception*, *31*(3), 315–321. https://doi.org/10.1068/p3129

Riley, K. G., & McGregor, K. K. (2012). Noise hampers children's expressive word learning. *Language, Speech, and Hearing Services in Schools*, *43*(3), 325–337. https://doi.org/10.1044/0161-1461(2012/11-0053)

Ristau, C. A., & Robbins, D. (1982). Language in the great apes: A critical review. *Advances in the Study of Behavior*, *12*, 141–255. https://doi.org/10.1016/S0065-3454(08)60048-0

Rodriguez, E. T., & Tamis-Lemonda, C. S. (2011). Trajectories of the home learning environment across the first 5years: Associations with children's vocabulary and literacy skills at prekindergarten. *Child Development*, *82*(4), 1058–1075. https://doi.org/10.1111/j.1467-8624.2011.01614.x

Rose, E. J., Simonotto, E., Spencer, E. P., & Ebmeier, K. P. (2006). The effects of escitalopram on working memory and brain activity in healthy adults during performance of the n-back task. *Psychopharmacology*, *185*(3), 339–347. https://doi.org/10.1007/s00213-006-0334-2

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, *133*(4), 2431–2443. https://doi.org/10.1121/1.4794379

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205. https://doi.org/10.1017/S0305000907008343

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–11928. https://doi.org/10.1126/science.274.5294.1926

Saito, A., & Shinozuka, K. (2013). Vocal recognition of owners by domestic cats (Felis catus). *Animal Cognition*, *16*(4), 685–690. https://doi.org/10.1007/s10071-013-0620-4

San Filippo, M. (2018). AVMA releases latest stats on pet ownership and veterinary care. Retrieved from https://www.avma.org/news/press-releases/avma-releases-latest-stats-pet-ownership-and-veterinary-care

Savage-Rumbaugh, E. S. (1986). *Ape language: From conditioned response to symbol*. Columbia University Press.

Schmidtke, J. (2016). The bilingual disadvantage in speech understanding in noise is likely a frequency effect related to reduced language exposure. *Frontiers in Psychology*, *7*, 678. https://doi.org/10.3389/fpsyg.2016.00678

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*(2), 1078–1083. https://doi.org/10.1121/1.428048

Shipley, K. G., & McAfee, J. G. (2015). *Assessment in speech-language pathology: A resource manual.* (5th ed.). Boston, MA: Cengage Learning.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, *3*(3), 365–394. https://doi.org/10.1207/S15327078IN0303_5

Siniscalchi, M., Lusito, R., Sasso, R., & Quaranta, A. (2012). Are temporal features crucial acoustic cues in dog vocal recognition? *Animal Cognition*, *15*(5), 815–821. https://doi.org/10.1007/s10071-012-0506-x

Sliwa, J., Duhamel, J. R., Pascalis, O., & Wirth, S. (2011). Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1735–1740. https://doi.org/10.1073/pnas.1008169108

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to improve hearing speech in noise: Biological mechanisms. *Cerebral Cortex*, *22*(5), 1180–1190. https://doi.org/10.1093/cercor/bhr196

Soproni, K., Miklósi, Á., Topál, J., & Csányi, V. (2001). Comprehension of human communicative signs in pet dogs (Canis familiaris). *Journal of Comparative Psychology*, *115*(2), 122–126. https://doi.org/10.1037//0735-7036.115.2.122

Springer, J. (2018). *The 2017-2018 APPA national pet owners survey debut*. Greenwich, CT.

Strain, G. M. (2011). *Deafness in dogs and cats*. Cambridge, MA: CABI.

Strain, G. M. (2012). Canine deafness. *Veterinary Clinics of North America - Small Animal Practice*, *42*(6), 1209–1224. https://doi.org/10.1016/j.cvsm.2012.08.010

Stuart, A. (2005). Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. *Ear and Hearing*, *26*(1), 78–88. https://doi.org/10.1097/00003446-200502000-00007

Swanson, H. L. (1983). A developmental study of vigilance in learning-disabled and nondisabled children. *Journal of Abnormal Child Psychology*, *11*(3), 415–429. https://doi.org/10.1007/BF00914249

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132. https://doi.org/10.1016/j.cogpsych.2004.06.001

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*(1), 16–34. https://doi.org/10.1016/j.jml.2006.07.002

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Thiessen, E. D., & Yee, M. N. (2010). Dogs, bogs, labs, and lads: What phonemic generalizations indicate about the nature of children's early word-form representations. *Child Development*, *81*(4), 1287–1303.

Tincoff, R., & Jusczyk, P. W. (2012). Six-Month-Olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics*, *67*(5), 867–875. https://doi.org/10.3758/BF03193539

Trehub, S. E., Bull, D., & Schneider, A. (1981). Infants' detection of speech in noise. *Journal of Speech and Hearing Research*, *24*, 202–206.

Trehub, S. E., Schneider, B. A., & Henderson, J. L. (1995). Gap detection in infants, children, and adults. *The Journal of the Acoustical Society of America*, *98*(5), 2532–2541. https://doi.org/10.1121/1.414396

van der Zee, E., Zulch, H., & Mills, D. (2012). Word generalization by a dog (Canis familiaris): is shape important? *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0049382

van Ooijen, B. (1996). Vowel mutability and lexical selection in English : *Memory & Cognition*, *24*(5), 573–583.

Vas, J., Topal, J., Pech, E., & Miklösi, Á. (2007). Measuring attention deficit and activity in dogs : A new application and validation of a human ADHD questionnaire. *Applied Animal Behaviour Science*, *103*, 105–117. https://doi.org/10.1016/j.applanim.2006.03.017

Vermeeren, A., Jackson, J. L., Muntjewerff, N. D., Quint, P. J., Harrison, E. M., & O'Hanlon, J. F. (1995). Comparison of acute alprazolam (0.25, 0.50 and 1.0 mg) effects versus those of Iorazepam 2 mg and placebo on memory in healthy volunteers using laboratory and telephone tests. *Psychopharmacology*, *118*, 1–9.

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development*, *81*(2), 517–527.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276. https://doi.org/10.1111/j.1467-7687.2004.00345.x

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Wascher, C. A. F., Szipl, G., Boeckle, M., & Wilkinson, A. (2012). You sound familiar: Carrion crows can differentiate between the calls of known and unknown heterospecifics. *Animal Cognition*, *15*(5), 1015–1019. https://doi.org/10.1007/s10071-012-0508-8

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289. https://doi.org/10.1037/0012-1649.34.6.1289

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*(1), 349–355. https://doi.org/10.1111/j.1467-8624.1981.tb03051.x

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63. https://doi.org/10.1016/S0163-6383(02)00093-0

Werner, L. A. (2007). Issues in human auditory development. *Journal of Communication Disorders*, *40*(4), 275–283. https://doi.org/10.1016/j.neuroimage.2013.08.045.The

Werner, L. A. (2013). Infants' detection and discrimination of sounds in modulated maskers. *The Journal of the Acoustical Society of America*, *133*(6), 4156–4167. https://doi.org/10.1121/1.4803903

Werner, L. A., & Bargones, J. Y. (1991). Sources of auditory masking in infants: Distraction effects. *Perception & Psychophysics*, *50*(5), 405–412. https://doi.org/10.3758/BF03205057

Werner, L. A., Marean, G. C., Halpin, C. F., Spetner, N. B., & Gillenwater, J. M. (1992). Infant auditory temporal acuity. *Society for Research in Child Development*, *63*(2), 260–272.

West, R. E., & Young, R. J. (2002). Do domestic dogs show any evidence of being able to count? *Animal Cognition*, *5*(3), 183–186. https://doi.org/10.1007/s10071-002-0140-0

Williams, K. N., & Perrott, D. R. (1972). Temporal resolution of tonal pulses. *The Journal of the Acoustical Society of America*, *51*(2B), 644–647.

Wobber, V., Hare, B., Koler-Matznick, J., Wrangham, R., & Tomasello, M. (2009). Breed differences in domestic dogs' (Canis familiaris) comprehension of human communicative signals. *Interaction Studies*, *10*(2), 206–224. https://doi.org/10.1075/is.10.2.06wob

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development*, *70*(1), 65–77. https://doi.org/10.1111/1467-8624.00006

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, *30*(4), 553–566. https://doi.org/10.1037/0012-1649.30.4.553

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750. https://doi.org/10.1038/358749a0

Xu, N., Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2013). Vowel hyperarticulation in parrot-, dog- and infant- directed speech. *Anthrozoös*, *26*(3), 373–380. https://doi.org/10.2752/175303713X13697429463592

Yip, M. J. (2006). The search for phonology in other species. *Trends in Cognitive Sciences*, *10*(10), 442–446. https://doi.org/10.1016/j.tics.2006.08.001

Yong, M. H., & Ruffman, T. (2015). Domestic dogs match human male voices to faces, but not for females. *Behaviour*, *152*(11), 1585–1600. https://doi.org/10.1163/1568539X-00003294